A short course on random matrices
Preliminary draft

Charles Bordenave

Preprint – Compiled June 17, 2014
Chapter 1

Why random matrices?

1.1 Some motivation (beyond curiosity)

1.1.1 In theoretical physics

Random matrix theory was born with the seminal work of Eugene Wigner in nuclear physics in the 50’s [Wig55, Wig58]. It still attracts a lot of attention in many branches of physics including quantum chaos, quantum gravity in two dimensions, . . .

1.1.2 In multivariate statistics

Initiated by Wishart [Wis28] in 1928. Consider \( X \in \mathbb{R}^p \) a centered Gaussian vector with covariance matrix \( \Sigma = \mathbb{E}XX^* \). Let \((X_i)_{i \geq 1}\) iid samples. Sample covariance matrix

\[
\Sigma_n = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^*
\]

For \( p \) fixed and \( n \) large : usual setting in multivariate statistic. Fundamental theorem of statistics, as \( n \to \infty \), a.s.

\[
\| \Sigma_n - \Sigma \| \to 0.
\]

where \( \| \cdot \| \) is a matrix norm. For example operator norm

\[
\|A\|_{2\to 2} = \max_{x \in \mathbb{R}^p} \frac{\| Ax \|_2}{\| x \|_2} = \max_{1 \leq i \leq n} \sqrt{\lambda_i(AA^*)}.
\]

**Random matrix regime** : what happens if \( p = p(n) \) grows to infinity with \( n \) ? (high dimensional data).

For example, assume for simplicity that \( \Sigma = I_p \) and \( p(n) \sim cn \), with \( 0 \leq c \leq 1 \), then, a.s.

\[
\| \Sigma_n - I_p \|_{2\to 2} \to (1 + \sqrt{c})^2 - 1.
\]

Also, as \( p \to \infty \), a.s.

\[
\log | \det(\Sigma_n - I_p) | = \sum_{i=1}^{p} \log | \lambda_i(\Sigma_n) - 1 | \to -\infty
\]

In the general case : is it possible to estimate the covariance \( \Sigma \) from \( \Sigma_n \) ?

See notably the monograph by Bai and Silverstein [BS10]. Applications in finance.
1.1.3 Numerical analysis


Spielman and Teng [ST04], Edelman and Rao [ER05]. Analysis of $X$ or $A + \sigma^2 X$ with $X$ random.

The condition number, for $A \in \mathcal{M}_{n,p}(\mathbb{R})$, $n \leq p$,

$$\kappa(A) = \sqrt{\lambda_{\max}(AA^*) / \lambda_{\min}(AA^*)}$$

plays a central role in the analysis of matrix algorithms.

1.1.4 In signal processing

Compressed sensing: $x_0 \in \mathbb{C}^n$, $A \in \mathcal{M}_{p \times n}(\mathbb{C})$, $p \ll n$.

$$y = Ax_0.$$

The system is undetermined. We suppose however that the signal $x_0$ is sparse: it contains a lot of 0. If $A$ are iid Gaussian variables with mean 0 and variance $1/n$, then, there exists $c > 0$ (Candès- Romberg-Tao [CRT06]) with high probability, for any vector $x_0$ with $|\text{supp}(x)| \leq cp/\log(n/p)$,

$$x_0 = \arg \min_{x:Ax=y} |\text{supp}(x)| = \arg \min_{x:Ax=y} \|x\|_1$$

($\ell^1$-recovery: linear programming works on the RHS). There is no chance that it could work with

$$\arg \min_{x:Ax=y} \|x\|_2.$$

Other available version with noise

$$y = Ax + b.$$

What are the other matrices $A$ which satisfy such amazing recovery property?

1.1.5 In wireless communication

Wireless channel: $b \in \mathbb{C}^n$, $s \in \mathbb{C}^p$, $A \in \mathcal{M}_{n \times p}(\mathbb{C})$.

$$r = As + b,$$

$b$ is a Gaussian noise vector with covariance $Ebb^* = \sigma^2 I_n$, $s$ is a sent signal. What is the Shannon capacity of such communication channel?

$$\max_{Q:Q^*Q \geq \frac{1}{\sigma^2} I_n, \text{tr}Q = 1} \log \det(\sigma^2 I_n + QA^*A).$$

For some relevant applications: $p$ and $n$ are large and that the channel itself is random: for example $A_{ij}$ iid (fading).

Another problem: how to effectively reconstruct the signal $s_1$ from the received vector $r$? Linear mean square error: $\hat{s}_1 = g^*r$. If $A = [x, X]$ then $g = (XX^* + \rho I_n)^{-1}x$. Assume $s_1$ mean 0, variance 1, covariance 0. Then the performance of the reconstruction is estimated by the signal to noise ratio:

$$\frac{E|g^*(s_1x)|^2}{E|g^*(r - s_1x)|^2} = E(xX^* + \sigma^2 I_n)^{-1}x^*.$$

See the monograph by Tulino and Verdu [TV04].
1.2. MAIN MATRIX MODELS

1.1.6 In other branches of mathematics
Connection (or conjectured connection) with analytic number theory, group theory, combinatorics, non-commutative geometry, in high-dimensional geometry . . .

1.2 Main matrix models

1.2.1 Wigner matrix
For $1 \leq i < j$

$$X_{ij} = \bar{X}_{ji}$$

iid with law $P$ on $\mathbb{C}$, independent of $X_{ii}, i \geq 1$ iid with common law $Q$ on $\mathbb{R}$. Then $X_n = (X_{ij})_{1 \leq i, j \leq n}$ is a random Hermitian matrix.

Important cases:
- $\sqrt{2} \Re (X_{ij}), \sqrt{2} \Im (X_{ij})$ and $X_{ii}$ are independent $N(0, 1)$: Gaussian Unitary Ensemble (GUE).
- $P = N(0, 1), Q = N(0, 2)$ (Var$(X_{11}) = 2$): Gaussian Orthogonal Ensemble (GOE).
- $P = Q$ is the Bernoulli law.

1.2.2 Wishart matrix
Also, very important (especially for applications) $(Y_{ij})_{i,j \geq 1}$ iid. We define the random matrix:

$$Y = (Y_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}.$$ 

Wishart matrix:

$$W = Y^*Y.$$ 

Important case: $\Re(Y_{ij})$ and $\Im(Y_{ij})$ independent $N(0, 1/2)$: the law of $W$ is called the Laguerre ensemble.

1.2.3 Random square matrix
$Y$ is as above with $n = p$.

Important case: $\Re(Y_{ij})$ and $\Im(Y_{ij})$ independent $N(0, 1/2)$: the law of $Y$ is then called the Ginibre ensemble.

1.2.4 Many other matrix models
Random unitary/orthogonal matrices, $Y_i = (Y_{ij}) \in \mathbb{R}^p$ sampled from a more general (non-product) law, random distance matrices, . . .

1.3 Some results in random matrix theory

1.3.1 Global regime

Convergence of ESD
If $A \in \mathcal{M}_n(\mathbb{C})$, eigenvalues $\lambda_1(A), \cdots, \lambda_n(A)$ counting multiplicities. We consider the empirical spectral distribution (ESD)

$$\mu_A = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(A)}$$
CHAPTER 1. WHY RANDOM MATRICES?

This is a global function of the spectrum. We consider the Wigner matrix $X_n$. The first fundamental result is Wigner’s semicircular law. It gives the convergence of the empirical spectral distribution.

To be more precise, we denote by $\Rightarrow$ the weak convergence of probability measures on a metric space $\mathcal{X}$: $\mu_n \Rightarrow \mu$ if for all bounded continuous functions $f : \mathcal{X} \to \mathbb{R}$,

$$\int f \, d\mu_n \to \int f \, d\mu.$$

If $Z_n$ has law $\mu_n$ and $Z$ has law $\mu$, we will then write $Z_n \overset{d}{\to} Z$ or $Z_n \overset{d}{\to} \mu$.

For example, if $(Z_i)_{i \geq n}$ are iid with common law $\mu$ then, a.s.

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i} \Rightarrow \mu.$$

Theorem 1.1 (Semi-circular Law). Assume that $\text{Var}(X_{12}) = 1$, then a.s.

$$\mu_{X/\sqrt{n}} \Rightarrow \mu_{sc}$$

where $\mu_{sc}(dx) = \frac{1}{2\pi} \sqrt{4-x^2} 1_{|x| \leq 2} \, dx$

We will give a few different proofs of this theorem. Note that only the second moment counts, not the first, neither all the others. This theorem is the analog of a law of large number.

For the case Wishart matrices, we have the Marcenko-Pastur’s Theorem [MP67].

Theorem 1.2 (Marcenko-Pastur Law). Assume that $\text{Var}(Y_{11}) = 1$, and $p/n \to c \in [0,1]$ then a.s.

$$\mu_{YY*/n} \Rightarrow \mu_{MP}$$

where $\mu_{MP}(dx) = \frac{1}{2\pi} \sqrt{(x-b_-)(b_+-x)} 1_{b_- \leq x \leq b_+} \, dx$ and $b_- = (1-\sqrt{c})^2$, $b_+ = (1+\sqrt{c})^2$.

The circular law deals with the spectrum of random non-Hermitian square matrices. First found by Metha 1967 [Meh67], Girko [Gir90], Bai [Bai97], . . . , Tao andVu [TV10a].

Theorem 1.3 (Circular Law). Assume that $\text{Var}(Y_{11}) = 1$, and $p = n$ then a.s.

$$\mu_{Y/\sqrt{n}} \Rightarrow \mu_c$$

where $\mu_c(dx) = \frac{1}{\pi} 1_{|x+iy| \leq 1} \, dx$.

Central Limit Theorem

If $(Z_i)$ is iid with law $\mu$, the central limit theorem asserts that for any function $f \in L^2(\mu)$,

$$\sqrt{n}\left(\int f \, d\mu_n - \int f \, d\mu\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(Z_i) - \mathbb{E} f(Z_i)) \overset{d}{\to} N(0, \text{Var}_\mu(f)).$$

In other words weakly $\sqrt{n}(\mu_n - \mu) \overset{d}{\to} W$ where $W$ is the Gaussian process (field) indexed by functions $f \in L^2(\mu)$ with covariance $\mathbb{E}[\langle W,f \rangle \langle W,g \rangle] = \int f \, gd\mu - \int f \, d\mu \int g \, d\mu$ (This is called the Brownian bridge).

For Wigner matrices, there is also an analog of this CLT but with a different rate! Let $f \in C^1_0(\mathbb{R})$. For GUE matrix, we have

$$n\left(\int f \, d\mu_{X/\sqrt{n}} - \int f \, d\mu_{sc}\right) = \sum_{i=1}^{n} f(\lambda_i) - n \int f(\lambda) \, d\mu_{sc} \overset{d}{\to} N(0,\sigma^2(f))$$
1.3. SOME RESULTS IN RANDOM MATRIX THEORY

where

\[ \sigma^2(f) = \frac{1}{4\pi^2} \int_{[-2,2]^2} \frac{(f(\lambda_1) - f(\lambda_2))^2}{(\lambda_1 - \lambda_2)^2} \frac{4 - \lambda_1 \lambda_2}{\sqrt{4 - \lambda_1^2} \sqrt{4 - \lambda_2^2}} d\lambda_1 d\lambda_2. \]

This result is also known for other Wigner matrices: the variance \( \sigma^2(f) \) depends on the fourth cumulant of the law of \( X_{12} \) (recall that \( \kappa_4(\mu) = \mathbb{E}[(X - \mathbb{E}X)^4] - 3\mathbb{E}[(X - \mathbb{E}X)^2]^2 \) where \( X \) has law \( \mu \)).

Also the CLT depends on the regularity of the function \( f \). For example if \( f = 1_I \) then \( \text{Var}(n \int f d\mu_{X/\sqrt{n}}) \sim c(\log n) \) (this is known for matrices of the GUE case and Wigner matrices with \( \kappa_4 = 0 \) and high enough moments, see Gustavsson [Gus05], Dallaporta and Vu [DV11]).

1.3.2 Local regime

We now consider a single or a few eigenvalues. If \( A \in \mathcal{H}_n \), \( \lambda_n(A) \leq \cdots \leq \lambda_1(A) \).

**Edge eigenvalues**

F"uredi and Komlós [FK81], Yin and Bai [BY88].

**Theorem 1.4** (Support convergence). Assume that \( \mathbb{E}X_{11} = \mathbb{E}X_{12} = 0 \), \( \text{Var}(X_{12}) = 1 \) and \( \mathbb{E}|X_{12}|^4, \mathbb{E}|X_{11}|^2 < \infty \) then a.s.

\[ \lim_{n} \lambda_1(X/\sqrt{n}) = - \lim_{n} \lambda_n(X/\sqrt{n}) = 2. \]

The conditions \( \mathbb{E}|X_{12}|^4, \mathbb{E}|X_{11}|^2 < \infty \) are necessary for the convergence to hold.

There are also fluctuation results, for the GUE it is known that

\[ n^{2/3}(\lambda_1(X/\sqrt{n}) - 2) \xrightarrow{d} TW_2. \]

The distribution \( TW_2 \) is not a usual probability measure from extremal value theory. It is the Tracy-Widom distribution, found in 1994, see [TW94]. In other cases, recent results due to Tao-Vu and Erdős-Schlein-Yau have extended this result to other Wigner matrices, see the surveys [Erd11], [TV12].

**Bulk eigenvalues**

Inside the spectrum, define the \( i \)-th \( n \)-quantile of \( \mu_{sc} \) as the element of \([-2, 2]\) defined by the formula

\[ \int_{-2}^{\bar{\lambda}_i(n)} f_{sc}(y)dy = \frac{i}{n}. \]

Wigner semicircle theorem asserts that if \(-2 < u < 2 \) and \( \bar{\lambda}_i(n) = u(1 + o(1)) \) then a.s.

\[ \lambda_i(X/\sqrt{n}) - \bar{\lambda}_i(n) \to 0. \]

In the GUE case we have, if \(-2 < u < 2 \) and \( \bar{\lambda}_i(n) = u(1 + o(1)) \) and

\[ \frac{n}{f_{sc}(u) \sqrt{\log n / 2\pi}} \left( \lambda_i(X/\sqrt{n}) - \bar{\lambda}_i(n) \right) \xrightarrow{d} N(0, 1) \]

(Costin and Lebowitz, [CL95], Soshnikov [Sos02]). Also known in some cases in the non-GUE case. Random eigenvalues are extremely rigid compared to independent variables.
CHAPTER 1. WHY RANDOM MATRICES?

Limit point process
More generally, we could consider the local regime and look at the point process, for \( u \in [-2, 2] \),
\[
\{ \lambda_i(\sqrt{n}X) - un, 1 \leq i \leq n \}
\]
the statistics of this process. For \(-2 < u < 2\), there is a limit point process which is described by a stationary determinantal point process with kernel sinus cardinal. \textit{xxx incomplete.}

1.3.3 Invertibility of random matrices and related results
Take \( n = p \) and consider the random square matrix \( Y \). Non-asymptotic estimates on
\[
\|Y\|_{2\to2}, \|Y^{-1}\|_{2\to2}
\]
For example assume that \( Y_{12} \) is a \( \pm 1 \) Bernoulli with parameter \( 1/2 \), there exists \( 0 < c < 1 \), such that
\[
\Pr( Y \text{ is singular } ) \leq c^n.
\]
Conjecture of Komlós, Kahn, and Szemerédi
\[
\Pr( Y \text{ is singular } ) = \left( \frac{1}{2} + o(1) \right)^n.
\]
Best result to date : Bourgain, Vu and Wood \cite{BVW10} with \( c = 1/\sqrt{2} + o(1) \).
Spielman and Teng conjecture
\[
\Pr( \sqrt{n}\|Y^{-1}\|_{2\to2} \leq t ) \leq t + c^n.
\]
with a constant in front of \( t \) : Rudelson and Vershynin \cite{RV08}, Rudelson \cite{Rud13}, with a limit in \( n \) on both side : Tao and Vu \cite{TV10b}.
For \( n \geq p \) and \( \E Y_{11} = 0 \), \( \E |Y_{11}|^2 = 1 \), explicit control on
\[
\left\| \frac{1}{n} Y^* Y - I_p \right\|_{2\to2}.
\]
Some mathematical issues of compressed sensing falls into this category, see Vershynin \cite{Ver12}.
Chapter 2

First approach to Wigner’s Theorem: the method of moments

We will give a first proof of Wigner’s Theorem.

2.1 Method of moments

Let \( Z \) be a real random variable with all its moments finite for all integer \( k \geq 1, \mathbb{E}[Z^k] = m_k < \infty \). Assume that there exists a unique probability measure \( P \) on \( \mathbb{R} \) such that for all integer \( k \geq 1, \int x^k dP = m_k \). From Carleman’s Theorem, this is indeed the case if

\[
\sum_{k \geq 1} m_{2k}^{-1} = \infty.
\]

If the random variable has bounded support, the Carleman condition is satisfied. Note that the Weierstrass theorem implies that a random variable with bounded support is uniquely determined by its moments.

Then, a commonly used method to prove that a sequence of real random variables \((Z_n)_{n \geq 1}\) converges weakly to a random variable \(Z\) is to show that for all integer \( k \geq 1, \lim_n \mathbb{E}[Z_n^k] = \mathbb{E}[Z^k] = m_k \).

Lemma 2.1. Assume that the law \( P \) is uniquely determined by its moments. If for all \( k \geq 1, \lim_n \int x^k dP_n(x) = \int x^k dP \) then \( P_n \Rightarrow P \).

Proof. We have \( \mathbb{E}Z_n^2 = m_2 + o(1) \). Hence, \( Z_n \) is tight and from Prohorov’s theorem \( \{\mathcal{L}(Z_n), n \geq 1\} \) is relatively compact. Let \( W \) be an accumulation point, since the law of \( Z \) is uniquely determined by its moments, it is sufficient to check that for any \( k \geq 1, \mathbb{E}W^k = m_k \). This amounts to prove that \( x \mapsto x^k \) is uniformly integrable for \((P_n)_{n \geq 1}\).

Let us check this by hand. Since \( \mathbb{E}Z_n^{2k} \) is uniformly bounded, we have \( \mathbb{P}(|Z_n| > t) \leq C t^{-2k} \) and from Portemanteau Theorem \( \mathbb{P}(|W| > t) \leq C t^{-2k} \). It follows that for any \( \varepsilon > 0 \), there is \( T \geq 1 \), such that \( \mathbb{E}|W|^k 1_{|W| \geq T} < \varepsilon \) and \( \mathbb{E}|Z_n|^k 1_{|Z_n| \geq T} < \varepsilon \). Consider \( f \) continuous \( f(x) = x^k \) on \([-T, T]\) and \( f(x) = \pm T \) on \( \mathbb{R} \setminus [-T, T] \). Then

\[
\mathbb{E}W^k = \mathbb{E}f(W) + \mathbb{E}W^k 1_{|W| \geq T} - \mathbb{E}f(W)^k 1_{|W| \geq T} = \mathbb{E}W^k 1_{|W| \geq T} - T^k \mathbb{P}(|W| \geq T) + \mathbb{E}f(W) - \mathbb{E}f(Z_n) + R,
\]

With \( |R| \leq 4 \varepsilon \).

There are many drawbacks to this method. First, the random variable \( Z_n \) needs to have finite moments of any order for all \( n \) large enough. Secondly, the computation of moments can be tedious. This method is however very robust.
2.2 Graphs, plane rooted tree and Catalan numbers

Let $G = (V, E)$ be a connected graph. A tree is a graph without cycles. Define the excess of $G$ as

$$\text{Exc}(G) = |E| - |V| - 1$$

**Lemma 2.2.** If $G$ is connected then $\text{Exc}(G) \geq 0$ and $\text{Exc}(G) = 0$ if and only if $G$ is a tree.

**Proof.** A possible proof by recursion. Otherwise, let $u \in V$ be a distinguished vertex and consider for all $v \in V \setminus \{u\}$ a shortest path $\{u_0(v), u_1(v), \ldots, u_k(v)\}$ from $v$ to $u : u_0(v) = v, u_k(v) = u$. Define the mapping $\sigma$ from $V \setminus \{u\}$ to $E$ by setting $\sigma(v) = \{v, u_1(v)\}$. Since the paths are the shortest possible, $\sigma$ is an injection, and it follows that $|V \setminus \{u\}| \leq |E|$. In the case of equality $|V \setminus \{u\}| = |E|$, $\sigma$ is a bijection. Note that $\sigma$ gives an orientation of the edges of $G$ and, by construction, any vertex in $V \setminus \{u\}$ has exactly one outgoing oriented edge and $u$ has no outgoing oriented edge. Observe also that path along oriented edges is a shortest path to $u$. In particular, there cannot be an oriented cycle and $G$ must be a tree. □

A rooted graph is a graph with a distinguished vertex. In a rooted tree $T = (V, E)$ rooted at $u$, the offsprings of $v \in V \setminus \{u\}$ are the set of vertices $w \in V \setminus \{u, v\}$ such that the shortest path from $w$ to $u$ is $(w, v, \ldots, u)$. A rooted plane tree is a tree where the offsprings of each vertex are ordered.

**Definition 2.3 (Depth-first search).** Let $T$ be a rooted plane tree. The depth-first search path in the unique connected closed path starting from the root which visits each edge twice and which preserves the orientation.

Note, that the depth-first search, gives a bijection $\varphi$ from the vertices $\{v_1, \ldots, v_\ell\}$ of a rooted plan tree into $\{1, \ldots, \ell\}$. The root is send to 1 and the vertices are ordered by order of appearance in the depth-first search. We say that two rooted plan trees $T$ and $T'$ are isomorphic if $\varphi(T) = \varphi(T')$ (where $\varphi$ acts transitively on the edges). We may then define $c_k$ as the number of (isomorphic classes) of rooted plane trees with $k$ edges. An isomorphism class of rooted plan tree is called an *unlabeled rooted plan tree*.

We have the recursion equation $c_0 = 1$ (convention) and for $k \geq 1$,

$$c_k = \sum_{\ell=0}^{k-1} c_\ell c_{k-\ell-1}.$$  \hfill (2.2.1)

(and thanks to the bijection with Dyck paths, $c_k \leq 2^{2^k}$). Now, define the generating function

$$S(z) = \sum_{k=0}^\infty c_k z^k$$

then radius $\geq 1/4$ and we find from (2.2.1)

$$S(z) = 1 + zS(z)^2.$$  \hfill (2.2.2)

Hence

$$S(z) = \frac{1 \pm \sqrt{1 - 4z}}{2z}.$$  

Since $S(0) = 1$, we find

$$S(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$  

By Taylor expansion,

$$\sqrt{1 - 4z} = 1 - 2z - \sum_{k=1}^\infty \frac{2^{-(k+1)}(2k - 1)(2k - 3)\cdots(1)}{(k + 1)!} (4z)^{k+1}.$$
2.3. MOMENTS OF SEMI-CIRCULAR LAW

Yielding,
\[ S(z) = 1 + 2 \sum_{k=1}^{\infty} \frac{2^{-(k+1)}(2k-1)(2k-3) \cdots (1)}{(k+1)!} (4z)^k \]

We find
\[ c_k = \frac{(2k)!}{(k+1)!k!} = \frac{(2k)!}{(k+1)!k!}. \]

2.3 Moments of semi-circular law

The moments of the semi-circular law are
\[ m_{2k+1} = 0 \]
and
\[ m_{2k} = c_k = \frac{(2k)!}{(k+1)!k!} = \frac{(2k)!}{(k+1)!k!}. \]

2.4 Computation of moments

We will follow the classical argument, as in [Gui09]. Recall that, for a random probability measure \( \mu \in \mathcal{P}(\mathcal{X}) \), we may define its expectation \( E\mu \in \mathcal{P}(\mathcal{X}) \), as for all Borel sets \( B \),
\[ (E\mu)(B) = E[\mu(B)]. \]

We will prove the following statement

**Proposition 2.4.** Assume \( E X_{11} = E X_{12} = 0, E X_{12}^2 = 1 \), \( X_{ij} \) real and \( X_{11}, X_{12} \) has finite moments of any order. Then
\[ E \mu X/\sqrt{n} \Rightarrow \mu_{sc}. \]

We set
\[ A = \frac{X}{\sqrt{n}}. \]

From the spectral theorem
\[ \int x^k d\mu_A = \frac{1}{n} \sum_{i=1}^{n} \lambda_i(A)^k = \frac{1}{n} E Tr A^k. \]

Hence, in view of lemma 2.1, it is sufficient to prove that

**Lemma 2.5.** For each integer \( k \),
\[ \frac{1}{n} E Tr A^k = m_k + O(1/\sqrt{n}). \]

**Proof.**
\[ \frac{1}{n} E Tr A^k = \frac{1}{n^{k/2+1}} \sum_{(i_1, \cdots, i_k)} \prod_{\ell=1}^{k} X_{i_\ell i_{\ell+1}} = \frac{1}{n^{k+1}} \sum_{(i_1, \cdots, i_k)} P(i), \]

where \( i_{k+1} = i_1 \)
and
\[ P(i) = E \prod_{\ell=1}^{k} X_{i_\ell i_{\ell+1}}. \]
12

CHAPTER 2. FIRST APPROACH TO WIGNER’S THEOREM: THE METHOD OF MOMENTS

Define \( G(i) = (V(i), E(i)) \) as the graph (with loops) obtained by setting \( V(i) = \{i_1, \cdots, i_k\} \) and \( E(i) = \{\{i_1, i_2\}, \cdots, \{i_k, i_1\}\} \) (by convention, \( x_j, 1 \leq j \leq J \) is a set : if \( x_1 = x_2 \), then \( \{x_1, x_2\} = \{x_1\} \)). For \( e \in E(i) \) we define the multiplicity of the edge \( e \) as

\[
m_e(i) = \sum_{\ell=1}^{k} \mathbf{1}(\{i_\ell, i_{\ell+1}\} = e).
\]

We fix \( i \). We note that \( G = G(i) \) is connected. Note also that, since \( \mathbb{E}X_{12} = 0 \), we may restrict ourselves to graphs such that for all \( e \in E \), \( m_e \geq 2 \) (otherwise \( P(i) = 0 \)). Therefore, the identity

\[
k = \sum_{e \in E} m_e
\]

yields

\[
|E| \leq \left\lceil \frac{k}{2} \right\rceil,
\]

where \( \lceil x \rceil \) is the integer part of \( x \). By lemma 2.2 we find

\[
|V| \leq |E| + 1 \leq \left\lceil \frac{k}{2} \right\rceil + 1.
\]

Also, by assumption and Hölder inequality,

\[
P(i) = \prod_{e \in E} (\mathbb{E}X_e^{m_e}) \quad \text{and} \quad |P(i)| \leq \prod_{e \in E} (\mathbb{E}|X_e|^{2k})^{\frac{m_e}{2k}} \leq \left( \mathbb{E}|X_e|^{2k} \right)^{\frac{1}{2}} = \beta_k.
\]

If \( \alpha_k \) is the number of graphs with vertex set \( \{1, \cdots, k\} \), we get that

\[
\left| \frac{1}{n} \mathbb{E} \text{Tr} A^k \right| \leq \alpha_k \beta_k n^{\frac{[k]}{2} + 1} n^{\frac{k}{2} + 1}.
\]

In particular, if \( k \) is odd the above expression is \( O(1/\sqrt{n}) \). If \( k = 2\ell \) is even then

\[
\frac{1}{n} \mathbb{E} \text{Tr} A^{2\ell} = \frac{1}{n^{\ell+1}} \left( \sum_{i \in \mathcal{I}(n, \ell)} 1 \right) + O \left( \frac{1}{n} \right),
\]

where \( \mathcal{I}(n, \ell) \) is the set of \( i = (i_1, \cdots, i_{2\ell}) \) such that \( G(i) \) is a tree with \( \ell \) edges and the multiplicity of each edge \( e \) has multiplicity exactly 2. From definition 2.3 the path \( i = (i_1, \cdots, i_{2\ell}) \) is a depth-first search of a rooted plane tree.

For \( i \in \mathcal{I}(n, \ell) \), we may map bijectively \( \{1, \cdots, \ell + 1\} \) into \( \{i_1, \cdots, i_{2\ell}\} \) by setting \( \sigma(1) = i_1 \), \( \sigma(2) = i_2 \) and \( \sigma(u) \) equal to the \( u \)-th distinct element in \( \{i_1, \cdots, i_{2\ell}\} \). Then \( T(i) = \sigma^{-1}(G(i)) \) is a rooted plan tree. Let \( \mathcal{T}(\ell) \) be the set of unlabelled rooted plane tree on \( \ell + 1 \) vertices. We consider the map \( \varphi : \mathcal{I}(n, \ell) \mapsto \mathcal{T}(\ell) \) which maps \( i \) to \( T(i) \). This map is surjective and for each element of \( T \in \mathcal{T}(\ell), |\varphi^{-1}(T)| = n \cdots (n - \ell) = n^{\ell+1}(1 + O(\frac{1}{n})) \). We find finally

\[
\frac{1}{n} \mathbb{E} \text{Tr} A^{2\ell} = \frac{1}{n^{\ell+1}} \left( \sum_{T \in \mathcal{T}(\ell)} n^{\ell+1}(1 + O(\frac{1}{n})) \right) + O \left( \frac{1}{n} \right) = |\mathcal{T}(\ell)| + O \left( \frac{1}{n} \right).
\]

The statement follows. \qed
2.4. COMPUTATION OF MOMENTS

2.4.1 Computation of joints moments

How to improve the above convergence of $\mu_A$ to a.s. ?

Proposition 2.6. Assume $\mathbb{E}X_{11} = \mathbb{E}X_{12} = 0$, $X_{ij}$ real, $\mathbb{E}X_{12}^2 = 1$ and $X_{12}$ has finite moments of any order. Then, a.s.

$$\mu_A \Rightarrow \mu_{sc}.$$

We could prove the following lemma

Lemma 2.7. For each integer $k$,

$$\text{Var}\left(\frac{1}{n} \text{Tr} A^k\right) = O(n^{-2}).$$

Then, it is the classical argument of the proof of the strong LLN with 4 moments. From the monotone convergence theorem

$$\mathbb{E} \sum_{n \geq 0} \left(\int x^k d\mu_A - \mathbb{E} \int x^k d\mu_A\right)^2 < \infty.$$

Hence a.s. $\sum_{n \geq 0} (\int x^k d\mu_A - \mathbb{E} \int x^k d\mu_A)^2 < \infty$ and, a.s. $\int x^k d\mu_A - \mathbb{E} \int x^k d\mu_A \to 0$. The proposition follows from lemma 2.1.

Proof of Lemma. We start with

$$\text{Var}\left(\frac{1}{n} \text{Tr} A^k\right) = \mathbb{E} \left(\frac{1}{n^{k/2+1}} \sum_{(i_1, \cdots, i_k)} \prod_{\ell=1}^k X_{i_{\ell+1}i_\ell} - P(i)\right)^2$$

$$= \frac{1}{n^{k+2}} \sum_{(i_1, \cdots, i_k), (j_1, \cdots, j_k)} P(i, j) - P(i)P(j),$$

where

$$P(i, j) = \mathbb{E} \prod_{\ell=1}^k X_{i_{\ell+1}i_\ell}X_{j_{\ell+1}j_\ell}.$$

We define $G(i, j) = (V(i, j), E(i, j), m(i, j))$ as the corresponding weighted graph. Note that $P(i, j) - P(i)P(j) = 0$ if $i$ and $j$ do not have two successive indices in common. Hence we may restrict to $G = G(i, j)$ connected. We have $\sum_{e \in E} m_e = 2k$ and $m_e \geq 2$. Hence $|E| \leq k$ and $|V| \leq |E| + 1 \leq k + 1$ and we get

$$\text{Var}\left(\frac{1}{n} \text{Tr} A^k\right) = O\left(\frac{n^{k+1}}{n^{k+2}}\right) = O\left(\frac{1}{n}\right).$$

Hence, we may restrict ourself to indices such that

$$|V| = |E| + 1 = k + 1 \text{ and } m_e = 2.$$

Up to changing the root, we can assume without loss of generality that $(i_1, i_2) = (j_1, j_2)$. Consider the path $\pi = (i_1, \cdots, i_k, i_1)$. Since $m_e = 2$, we have $i_k \neq i_2$. Hence $\pi$ is a closed path in $G$ which contains a cycle. This contradicts the assumption that $G$ is a tree. Therefore, since does not occur, we have $|E| \leq k - 1 \text{ and } |V| \leq k$. It follows that

$$\text{Var}\left(\frac{1}{n} \text{Tr} A^k\right) = O\left(\frac{n^k}{n^{k+2}}\right) = O\left(\frac{1}{n^2}\right).$$

It concludes the proof.

Remark 2.8. From here, we could also prove a CLT for the moments. See [AGZ10, theorem 2.1.31].

Remark 2.9. Considering oriented graphs instead of graphs, the same proof can be adapted for the Hermitian case.
Chapter 3

Perturbation of matrices and application to concentration inequalities

We will need more tools to develop the theory.

3.1 Fundamental matrix inequalities

3.1.1 Variational formula for the eigenvalues

We order the eigenvalues of $A \in \mathcal{H}_n$ non-increasingly

$$\lambda_n(A) \leq \cdots \leq \lambda_1(A).$$

**Lemma 3.1** (Courant-Fischer min-max theorem). Let $A \in \mathcal{H}_n(\mathbb{C})$. Then

$$\lambda_k(A) = \max_{H: \dim(H) = k} \min_{x \in H, \|x\|_2 = 1} \langle Ax, x \rangle.$$

**Proof.** Let $u_i$ be an eigenvector basis of $A$ associated to $\lambda_1, \cdots, \lambda_n$. We choose $H = \text{span}(u_1, \cdots, u_k)$. We find

$$\max_{H: \dim(H) = k} \min_{x \in H, \|x\|_2 = 1} \langle Ax, x \rangle \geq \lambda_k.$$

On the other hand, let $H$ be a vector space such that $\dim(H) = k$. Define $S = \text{span}(u_{n+1}, \cdots, u_k)$ so that $\dim(S) = n - k + 1$. Since

$$n \geq \dim(H \cup S) = \dim(H) + \dim(S) - \dim(S \cap H)$$

we find $S \cap H \neq 0$. In particular,

$$\min_{x \in H, \|x\|_2 = 1} \langle Ax, x \rangle \leq \lambda_k.$$

\hfill \Box

3.1.2 Interlacing of eigenvalues

An important corollary of the Courant-Fischer min-max theorem is the interlacing of eigenvalues. By convention if $A \in \mathcal{H}_n(\mathbb{C})$, we set for integer $i \geq 1$,

$$\lambda_{n+i}(A) = -\infty \quad \text{and} \quad \lambda_{1-i}(A) = +\infty \quad (3.1.1)$$
Lemma 3.2 (Weak interlacing). Let $A, B$ in $\mathcal{H}_n(\mathbb{C})$ and assume that $\dim(A - B) = r$ Then, for any $1 \leq k \leq n$,
\[ \lambda_{k+r}(A) \leq \lambda_k(B) \leq \lambda_{k-r}(A). \]

**Proof.** We prove $\lambda_{k+r}(A) \leq \lambda_k(B)$. We may assume that $k + r \leq n$. By definition, for some vector space $H$ of dimension $k + r$,
\[ \lambda_{k+r}(A) = \min_{x \in H, \|x\|_2 = 1} \langle Ax, x \rangle. \]
Take $H' = H \cap \ker(E)$, where $E = A - B$. By construction
\[ \lambda_{k+r}(A) \leq \min_{x \in H', \|x\|_2 = 1} \langle Ax, x \rangle = \min_{x \in H', \|x\|_2 = 1} \langle Bx, x \rangle \leq \lambda_{k'}(B) \]
where $k' = \dim(H')$. Now, the inequality,
\[ n - \dim(H') \leq (n - \dim(H)) + \dim(\text{im}(E)) \]
yields $k' \geq k$. This concludes the proof of the inequality $\lambda_{k+r}(A) \leq \lambda_k(B)$. For the proof of $\lambda_k(B) \leq \lambda_{k-r}(A)$, we may assume that $k - r \geq 1$. Then, simply replace $A$ and $B$ in the above argument. 

There are variants of the above interlacing inequality. The following can be proved as the above lemma.

Lemma 3.3 (Strong Interlacing). Let $A, B$ in $\mathcal{H}_n$ and assume that $A = B + \rho xx^*$ with $\|x\|_2 = 1$ and $\rho \geq 0$. Then
\[ \lambda_n(B) \leq \lambda_n(A) \leq \lambda_{n-1}(B) \leq \cdots \leq \lambda_2(A) \leq \lambda_1(B) \leq \lambda_1(A). \]

Lemma 3.4 (Cauchy law for minor interlacing). Let $A, B$ in $\mathcal{H}_n$ and assume that $B$ is a principal minor of $A$. Then
\[ \lambda_n(A) \leq \lambda_{n-1}(B) \leq \cdots \leq \lambda_2(A) \leq \lambda_1(B) \leq \lambda_1(A). \]

We now give a perturbation inequality which is a consequence of interlacing. For $\mu, \mu'$ two real probability measure, we introduce the Kolomogorov-Smirnov distance
\[ d_{KS}(\mu, \mu') = \sup_{t \in \mathbb{R}} \|\mu(-\infty, t) - \mu'(-\infty, t)\| = \|F_\mu - F_{\mu'}\|_\infty. \]
The Kolomogorov-Smirnov distance is closely related to functions with bounded variations. More precisely, for $f : \mathbb{R} \to \mathbb{R}$ the bounded variation distance is defined as
\[ \|f\|_{BV} = \sup_{k \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |f(x_{k+1}) - f(x_k)|, \]
where the supremum is over all sequence $(x_k)_{k \in \mathbb{Z}}$ with $x_n \leq x_{n+1}$. If the $f = 1((-\infty, t))$ then $\|f\|_{BV} = 1$ while if the derivative of $f$ is in $L^1(\mathbb{R})$, we have
\[ \|f\|_{BV} = \int |f'| \]
We have a variational formula for the Kolomogorov-Smirnov distance.
\[ d_{KS}(\mu, \mu') = \sup \left\{ \int fd\mu - \int f d\mu' : \|f\|_{BV} \leq 1 \right\}. \]
Choosing $f = 1((−∞, t))$ gives the inequality $d_{KS} ≥ ⋯$. The other way around, if $f ∈ C_0^1(\mathbb{R})$ and $τ$ is a continuity point of $F_μ$, we have the integration by part formula
\[
\int_{−∞}^{τ} f(t)dμ(t) = f(τ)F_μ(τ) − \int_{−∞}^{τ} f'(t)F_μ(t)dt.
\]
This yields, letting $τ$ tend to infinity,
\[
\int f(t)dμ(t) − \int f(t)dμ'(t) = − \int f'(t)(F_μ(t) − F_μ'(t))dt.
\]
In particular,
\[
\left| \int f dμ − \int f dμ' \right| ≤ \int |f'(t)||F_μ(t) − F_μ'(t)| dt = \|f\|_{BV} \|F_μ − F_μ'\|_∞.
\] (3.1.2)

Using the density of $C_0^1$ for $\|\cdot\|_{BV}$ norm, this inequality can also be extended to any $f$ with $\|f\|_{BV} < ∞$. We will apply it when $μ$ and $μ'$ are empirical probability measures where it is easy to check from the definition. We have the following consequence of interlacing.

**Lemma 3.5** (Rank inequality for ESD). *Let $A, B$ in $H_n(\mathbb{C})$ and assume that $\dim(A − B) = r$. Then,
\[
d_{KS}(μ_A, μ_B) ≤ \frac{r}{n},
\]*
and for any $f$ with $\|f\|_{BV} < ∞$,
\[
\left| \int f(t)dμ_A(t) − \int f(t)dμ_B(t) \right| ≤ \left( \frac{r}{n} \right) \|f\|_{BV}.
\]

**Proof.** Fix $t ∈ \mathbb{R}$. Let $k$ and $k'$ be the smallest indices such that $λ_k(A) ≤ t$ and $λ_{k'}(B) < t$ (recall our convention (3.1.1)). By lemma 3.2, we find
\[
|k − k'| ≤ r.
\]
This yields
\[
|F_μ(t) − F_μ'(t)| = \left| \frac{(n + 1 − k) − (n + 1 − k')}{n} \right| ≤ \frac{r}{n}.
\]

This gives the first statement. The second statements follows from (3.1.2). \[\square\]

### 3.1.3 Hoffman-Wielandt inequality

We now present another matrix inequality which is particularly useful.

**Lemma 3.6** (Hoffman-Wielandt inequality). *Let $A, B$ in $H_n(\mathbb{C})$,
\[
\sum_{i=1}^{n} |λ_i(A) − λ_i(B)|^2 ≤ Tr(A − B)^2 = \|A − B\|_F^2.
\]

**Proof.** Proof in [AGZ10]. \[\square\]

For $p ≥ 1$, $μ$, $μ'$ two real probability measure such that $∫ |x|^p dμ$ and $∫ |x|^p dμ'$ are finite. We define the $L^p$-Wasserstein distance as
\[
W_p(μ, μ') = \left( \inf \int_{\mathbb{R} \times \mathbb{R}} |x − y|^p dπ \right)^{\frac{1}{p}}
\]
where the infimum is over all coupling \( \pi \) of \( \mu \) and \( \mu' \) (i.e. \( \pi \) is probability measure on \( \mathbb{R} \times \mathbb{R} \) whose first marginal is equal to \( \mu \) and second marginal is equal to \( \mu' \)). Note that Hölder inequality gives for \( 1 \leq p \leq p' \),

\[
W_p \leq W_{p'}.
\]

For any \( p \geq 1 \), if \( W_p(\mu_n, \mu) \) converges to 0 then \( \mu_n \rightharpoonup \mu \). This follows for example from Kantorovich-Rubinstein duality

\[
W_1(\mu, \mu') = \sup \left\{ \int f \, d\mu - \int f \, d\mu' : \|f\|_{L^1} \leq 1 \right\}.
\]

**Corollary 3.7** (Hoffman-Wielandt inequality for ESD). Let \( A, B \) in \( \mathcal{H}_n(\mathbb{C}) \), then

\[
W_2(\mu_A, \mu_B) \leq \sqrt{\frac{1}{n} \text{Tr}(A - B)^2}.
\]

**Proof.** Consider the coupling \( \pi \) of \( (\mu_A, \mu_B) \) defined as

\[
\pi = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A), \lambda_i(B)}.
\]

We find

\[
\int_{\mathbb{R} \times \mathbb{R}} |x - y|^2 \, d\pi \leq \frac{1}{n} \text{Tr}(A - B)^2.
\]

The left hand side is lower bounded \( W_2^2(\mu_A, \mu_B) \) by construction (in fact it is even equal).

In the next corollary, we identify \( \mathbb{C} \) and \( \mathbb{R}^2 \).

**Corollary 3.8** (Continuity of the spectrum). Let \( f : \mathbb{R} \to \mathbb{R} \to \mathbb{R} \) be a Lipschitz function. Then the map \( F : \mathbb{R}^{n^2} \to \mathbb{R} \)

\[
F : (X_{ij})_{1 \leq i \leq j \leq n} \to f(\lambda_1(X), \ldots, \lambda_n(X)).
\]

is Lipschitz with constant \( \sqrt{2} \|f\|_{L^1} \). In particular, if \( g : \mathbb{R} \to \mathbb{R} \) is Lipschitz,

\[
G : (X_{ij})_{1 \leq i \leq j \leq n} \to \frac{1}{n} \text{Tr}g(X/\sqrt{n}) = \int gd\mu_{X/\sqrt{n}}
\]

is Lipschitz with constant \( \sqrt{2} \|g\|_{L^1}/n \).

**Proof.** From Hoffman-Wielandt inequality, we get

\[
|F(X) - F(Y)|^2 \leq \|f\|_{L^1}^2 \sum_{k=1}^{n} |\lambda_k(X) - \lambda_k(Y)|^2 \leq \|f\|_{L^1}^2 \text{Tr}(X - Y)^2.
\]

Now, construction,

\[
\text{Tr}(X - Y)^2 = \sum_{i=1}^{n} |X_{ii} - Y_{ii}|^2 + 2 \sum_{i>j} |X_{ij} - Y_{ij}|^2 \leq 2 \|X - Y\|_2^2.
\]

The second statement follows from the first statement, since the map

\[
(x_1, \ldots, x_n) \to \sum_{i=1}^{n} x_i
\]

is Lipschitz with constant \( \sqrt{n} \) (thanks to Cauchy-Schwartz inequality).
3.2 General Wigner’s semi-circular law

Proposition 3.9. Let $X_{ij}$ be a real Wigner matrix such that $X_{12}$ has law $P$ and and $X_{11}$ has law $Q$. If $\text{Var}(X_{12}) = 1$, then with a.s.

$$\mu_{X/\sqrt{n}} \sim \mu_{sc}.$$  

For the complex case, see Bai and Silverstein ”Spectral Analysis of Large Dimensional Random Matrices” §2.1.3

Proof. We define the distance

$$d(\mu, \mu') = \sup \left\{ \int f d\mu - \int f d\mu' : \|f\|_L \leq 1, \|f\|_{BV} \leq 1 \right\},$$

so that we can apply both corollary [3.7] and lemma [3.5] to this distance :

$$d(\mu_A, \mu_B) \leq \min \left( \frac{\text{rank}(A - B)}{n}, \sqrt{\frac{1}{n} \text{Tr}(A - B)^2} \right).$$

The distance $d$ is a metric for the weak convergence. Hence, in order to prove the theorem, it is sufficient to prove that, for any $\varepsilon > 0$, a.s.

$$\limsup_n d(\mu_{X/\sqrt{n}}, \mu_n) \leq 3\varepsilon,$$

for some sequence of probability measures such that a.s. $\mu_n \sim \mu_{sc}$. We first truncate the off-diagonal term, let $\varepsilon > 0$. For some $\tau > 0$ we set

$$X_{ij}^{(1)} = \begin{cases} X_{ij}1(|X_{ij}| \leq \tau) & \text{if } i \neq j \\ X_{ii} & \text{otherwise} \end{cases}$$

Then from Corollary, [3.7]

$$d(\mu_{X/\sqrt{n}}, \mu_{X^{(1)}/\sqrt{n}}) \leq \sqrt{\frac{1}{n^2} \sum_{i \neq j=1}^n |X_{ij}^2| 1(|X_{ij}| > \tau)}.$$  

By taking $\tau$ large enough, we have

$$E|X_{12}^2|1(|X_{12}| > \tau) = \int_\tau^\infty x^2 dP < \varepsilon^2.$$  

From the strong law of large numbers, we obtain. a.s.

$$\limsup_n d(\mu_{X/\sqrt{n}}, \mu_{X^{(1)}/\sqrt{n}}) \leq \varepsilon.$$  

We may also have chosen $\tau$ large enough so that

$$\sigma^2 = \text{Var}(X_{12}1(|X_{ij}| \leq \tau))$$

satisfies

$$\left| 1 - \frac{1}{\sigma^2} \right| \leq \varepsilon.$$  

Then we center the off-diagonal term and consider the matrix, with $m = E X_{12}1(|X_{ij}| \leq \tau),$

$$X^{(2)} = X^{(1)} - mJ.$$
Since $J$ has rank 1, from lemma 3.5 we find
\[ d(\mu_{X^{(2)}/\sqrt{n}}, \mu_{X^{(3)}/\sqrt{n}}) \leq \frac{1}{n}. \]

We now treat the diagonal terms.
\[
X_{ij}^{(3)} = \begin{cases} 
X_{ij}^{(2)} & \text{if } i \neq j \\
X_{ii}1(|X_{ii}| \leq \tau) & \text{otherwise}
\end{cases}
\]

Then from lemma 3.5 we find
\[
d(\mu_{X^{(2)}/\sqrt{n}}, \mu_{X^{(3)}/\sqrt{n}}) \leq \frac{1}{n} \sum_{i=1}^{n} 1(|X_{ii}| > \tau).
\]

By taking $\tau$ large enough, we have
\[
P(|X_{11}| > \tau) = \int_{\tau}^{\infty} dQ < \varepsilon.
\]

From the strong law of large numbers, we obtain, a.s.
\[
\limsup_{n} d(\mu_{X^{(2)}/\sqrt{n}}, \mu_{X^{(3)}/\sqrt{n}}) \leq \varepsilon.
\]

Finally, introduce the matrix
\[
X_{ij}^{(4)} = \begin{cases} 
X_{ij}^{(3)} & \text{if } i \neq j \\
0 & \text{otherwise}
\end{cases}
\]

Then, from Corollary 3.7 we find
\[
d(\mu_{X^{(4)}/\sqrt{n}}, \mu_{X^{(3)}/\sqrt{n}}) \leq \sqrt{\frac{1}{n^2} \sum_{i=1}^{n} |X_{ij}|^2 1(|X_{ii}| \leq \tau) \leq \frac{\tau}{\sqrt{n}},}
\]

which goes to 0. Let $m = E X_{12}$. We notice that the matrix
\[
Y = X^{(4)}/\sigma,
\]
is a Wigner matrix which satisfies the hypothesis of Proposition 2.6. In particular, a.s.,
\[
\mu_{Y/\sqrt{n}} \rightsquigarrow \mu_{sc}
\]
while, we certainly have,
\[
d(\mu_{X^{(4)}/\sigma \sqrt{n}}, \mu_{Y/\sqrt{n}}) \leq \left| 1 - \frac{1}{\sigma} \right| \leq \varepsilon.
\]

This concludes the proof.

\[\square\]

### 3.3 Concentration of measure for random matrices

#### 3.3.1 With Bounded Martingale difference inequality

**Theorem 3.10** (Concentration of ESD with independent half-rows). Let $M \in \mathcal{H}_n(\mathbb{C})$ be an Hermitian random matrix and for $1 \leq k \leq n$, define the variables $M_k = (M_{kj})_{1 \leq j \leq k} \in \mathbb{C}^k$. If the variables $(M_k)_{1 \leq k \leq n}$ are independent, then for any $f : \mathbb{R} \to \mathbb{R}$ such that $\|f\|_{BV} \leq 1$ and every $t \geq 0$,
\[
P \left( \left| \int f d\mu_M - E \int f d\mu_M \right| \geq t \right) \leq 2 \exp \left( -\frac{nt^2}{8} \right).
\]
3.3. CONCENTRATION OF MEASURE FOR RANDOM MATRICES

The proof is based on Azuma-Hoeffding’s inequality. Let \( X_1 \cdots X_n \) be metric spaces and let \( F \) be a measurable function on \( X = X_1 \times \cdots \times X_n \) and \( P \) a product measure on \( X \). There is very powerful tool to bound the deviation of \( F \) from its mean when \( F \) is Lipschitz for a weighted Hamming pseudo-distance, i.e. for every \( x \) and \( y \) in \( X 

\begin{equation}
|F(x) - F(y)| \leq \sum_{k=1}^{n} c_k 1_{x_k \neq y_k}. \tag{3.3.1}
\end{equation}

for some \( c = (c_1, \ldots, c_n) \in \mathbb{R}_+^n \). We denote by \( \|y\|_2 = \sqrt{\sum_i y_i^2} \), the usual Euclidean norm.

**Theorem 3.11** (Azuma-Hoeffding’s inequality). Let \( F \) be as above, then

\[
P \left( F - \int F dP \geq t \right) \leq \exp \left( -\frac{t^2}{2\|c\|_2^2} \right).
\]

This type of result is called a concentration inequality. It has found numerous applications in mathematics over the last decades. For more on concentration inequalities, we refer to [Led01].

As a corollary, we deduce the Hoeffding’s inequality.

**Corollary 3.12** (Hoeffding’s inequality). Let \((X_k)_{1 \leq k \leq n}\) be an independent sequence of real random variables such that for all integer \( k \), \( X_k \in [a_k, b_k] \). Then,

\[
P \left( \sum_{k=1}^{n} X_k - EX_k \geq t \right) \leq \exp \left( -\frac{t^2}{2\sum_{k=1}^{n} (b_k - a_k)^2} \right). \tag{3.3.2}
\]

The proof of Theorem 3.11 will be based on a lemma due to Hoeffding.

**Lemma 3.13.** Let \( X \) be real random variable in \([a, b]\) such that \( \mathbb{E}X = 0 \). Then, for all \( \lambda \geq 0 \),

\[
\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2(b-a)^2}{8}}.
\]

**Proof.** By the convexity of the exponential,

\[
e^{\lambda X} \leq \frac{b - X}{b - a} e^{\lambda a} + \frac{X - a}{b - a} e^{\lambda b}.
\]

Taking expectation, we obtain, with \( p = -a/(b - a) \),

\[
\mathbb{E}e^{\lambda X} \leq \frac{b}{b - a} e^{\lambda a} - \frac{a}{b - a} e^{\lambda b} = \left( 1 - p + pe^{\lambda(b-a)} \right) e^{-p\lambda(b-a)} = e^{\varphi(\lambda(b-a))},
\]

where \( \varphi(x) = -px + \ln(1 - p + pe^x) \). The derivatives of \( \varphi \) are

\[
\varphi'(x) = -p + \frac{pe^x}{(1 - p)e^{-x} + p} \quad \text{and} \quad \varphi''(x) = \frac{p(1 - p)}{((1 - p)e^{-x} + p)^2} \leq \frac{1}{4}.
\]

Since \( \varphi(0) = \varphi'(0) = 0 \), we deduce from Taylor expansion that

\[
\varphi(x) \leq \varphi(0) + x\varphi'(0) + \frac{x^2}{2} \varphi''(\infty) \leq \frac{x^2}{8}.
\]
Proof of Theorem 3.11} Let \( (X_1, \ldots, X_n) \) be a random variable on \( \mathcal{X} \) with distribution \( P \). We shall prove that
\[
P(F(X_1, \ldots, X_n) - \mathbb{E}F(X_1, \ldots, X_n) \geq t) \leq \exp \left( -\frac{t^2}{2\|c\|^2_2} \right).
\]

For integer \( 1 \leq k \leq n \), let \( F_k = \sigma(X_1, \ldots, X_k) \), \( Z_0 = \mathbb{E}F(X_1, \ldots, X_n) \), \( Z_k = \mathbb{E}[F(X_1, \ldots, X_n|F_k)] \), \( Z_n = F(X_1, \ldots, X_n) \). We also define \( Y_k = Z_k - Z_{k-1} \), so that \( \mathbb{E}[Y_k|F_{k-1}] = 0 \). Finally, let \( (X'_1, \ldots, X'_n) \) be an independent copy of \( (X_1, \ldots, X_n) \). If \( \mathbb{E}' \) denote the expectation over \( (X'_1, \ldots, X'_n) \), we have
\[
Z_k = \mathbb{E}'F(X_1, \ldots, X_k, X'_{k+1}, \ldots, X'_n).
\]
It follows by \( \text{(3.3.1)} \)
\[
Y_k = \mathbb{E}'F(X_1, \ldots, X_k, X'_{k+1}, \ldots, X'_n) - \mathbb{E}'F(X_1, \ldots, X_k-1, X'_{k-1}, \ldots, X'_n) \in [-c_k, c_k].
\]

Since \( \mathbb{E}[Y_k|F_{k-1}] = 0 \), we may apply Lemma 3.13 for every \( \lambda \geq 0 \),
\[
\mathbb{E}[e^{\lambda Y_k}|F_{k-1}] \leq e^{\frac{\lambda^2 c_k^2}{2}}.
\]
This estimates does not depend on \( F_{k-1} \), it follows that
\[
\mathbb{E}e^{\lambda(Z_n-Z_0)} = \mathbb{E}[e^{\lambda \sum_{k=1}^n Y_k}] \leq e^{\frac{\lambda^2 \|c\|^2_2}{2}}.
\]
From Chernov bound, for every \( \lambda \geq 0 \),
\[
P(F(X_1, \ldots, X_n) - \mathbb{E}F(X_1, \ldots, X_n) \geq t) \leq \exp \left( -\lambda t + \frac{\lambda^2 \|c\|^2_2}{2} \right).
\]
Optimizing over the choice of \( \lambda \), we choose \( \lambda = t/\|c\|^2_2 \).

Proof. For any \( x = (x_1, \ldots, x_n) \in \mathcal{X} := \{ (x_i)_{1 \leq i \leq n} : x_i \in \mathbb{C}^{i-1} \times \mathbb{R} \} \), let \( H(x) \) be the \( n \times n \) Hermitian matrix given by \( H(x)_{ij} = x_{ij} \) for \( 1 \leq j \leq i \leq n \). We thus have \( M = H(M_1, \ldots, M_n) \). For all \( x \in \mathcal{X} \) and \( x' \in \mathbb{C}^{i-1} \times \mathbb{R} \), the matrix
\[
H(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - H(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n)
\]
has only the \( i \)-th row and column possibly different from 0, and thus
\[
\text{rank} \left( H(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - H(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n) \right) \leq 2.
\]
Therefore from lemma 3.3 we obtain,
\[
\left| \int f(\lambda)d\mu_{H(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)} - \int f(\lambda)d\mu_{H(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n)} \right| \leq 2n^{-1}.
\]
The desired result follows now from the Azuma–Hoeffding inequality.

\[ \square \]

### 3.3.2 With Logarithmic Sobolev inequality

We are now going to derive optimal concentration inequalities. We follow Section 2.3.2 in \[\text{ACZ10} \].

**Definition 3.14** (Logarithmic Sobolev inequality (LSI)). A probability measure \( P \) on \( \mathbb{R}^n \) satisfies LSI with constant \( c \) if for any differentiable function \( f \in L^2(P) \)
\[
\text{Ent}_P(f^2) = \int f^2 \log \left( \frac{f^2}{\int f^2dP} \right) dP \leq c \int \| \nabla f \|_2^2 dP = c \mathcal{E}_P(\nabla f),
\]
where \( \| \nabla f \|_2^2 = \sum_{i=1}^n (\partial_i f)^2 \).
3.3. CONCENTRATION OF MEASURE FOR RANDOM MATRICES

Recall that the entropy bound which is naturally related to transport inequalities. For example, the Pinsker's inequality relates the entropy to the total variation distance between two probability measures:

\[ d_{TV}(\mu, \mu') = \sup \| \mu(A) - \mu(A') \| \leq \sqrt{2 \text{Ent}_\mu \left( \frac{\partial \mu'}{\partial \mu} \right)}. \]

The definition of LSI is due to Léonard Gross 1975. It bounds the entropy by an energy. It is closely related to hypercontractivity in semi-group theory. Refer to Ané et al. [ABC+00], to Ledoux [Led01]. For techniques to prove LSI, see also Villani Chap. 21-22 "optimal transport, old and new" and Guionnet and Zegarlinski [GZ03]. With the proper notion energy \( E_P(f) \), the definition of LSI extends in a context more general than \( \mathbb{R}^n \).

For us, it will be important that the standard Gaussian on \( \mathbb{R} \) satisfies LSI(2). More generally, let \( V : \mathbb{R}^n \to \mathbb{R} \cup \{ \infty \} \) such that \( V(x) - c||x||^2_2/4 \) is convex. Then the probability \( P(dx) = Z^{-1}e^{V(x)}dx \) satisfies LSI(c), it is a consequence of Bakry-Émery criterion, see Bobkov-Ledoux 2000. For the record, on space with two points \( \{0,1\} \), setting \( E_P(f) = P(0)P(1)|f(0) - f(1)|^2 \), the Bernoulli law on \( \{0,1\} \) with parameter \( p \) satisfies LSI(c(p)), with \( c(1/2) = 2 \) and

\[ c(p) = \frac{\log(q) - \log(p)}{q - p}. \]

Lemma 3.15 (Properties of LSI).

- Homogeneity : if \( P \) satisfies \( \text{LSI}(c) \) and \( X \overset{d}{=} P \), then the law of \( \sigma X + x \) satisfies \( \text{LSI}(\sigma^2 c) \).

- Tensorization : for \( i \in \{1,2\} \), if \( P_i \) is a probability measure on \( \mathbb{R}^{m_i} \) which satisfies satisfies \( \text{LSI}(c_i) \), then \( P_1 \otimes P_2 \) satisfies \( \text{LSI}(\max(c_1, c_2)) \).

Proof. Only the tensorization property deserves a proof. We write \( P(f) \) in place of \( E_P(f) \). We decompose the entropy

\[
\text{Ent}_{P_1 \otimes P_2}(f^2) = P_1 \otimes P_2 \left( f^2 \log \frac{f^2}{P_1 \otimes P_2(f^2)} \right)
= P_2 \left( P_1 \left( f^2 \log \frac{f^2}{P_1 f^2} \right) \right) + P_2 \left( P_1(f^2) \log \frac{P_1 f^2}{P_1 \otimes P_2(f^2)} \right)
= P_2 \left( \text{Ent}_{P_1}(f^2) \right) + \text{Ent}_{P_2}(P_1 f^2).
\]

Recall the variational formula for the entropy :

\[ \text{Ent}_P(h) = \sup \{ P(hg) : P(e^g) = 1 \}, \]

(which follows from the inequality : \( xy \leq x \log x - x + e^y \), for any \( x > 0, y \in \mathbb{R} \) which is applied to \( x = h/Ph \) and \( y = g \)). This yields to, for any function \( g : \mathbb{R}^{m_2} \to \mathbb{R} \) with \( E_{P_2} e^g = 1 \),

\[ P_2 \left( P_1 \left( f^2 g \right) \right) = P_1 \left( P_2 \left( f^2 g \right) \right) \leq P_1 \left( \text{Ent}_{P_2}(f^2) \right). \]

Taking the supremum over all \( g \), yields the tensorization inequality for the entropy :

\[ \text{Ent}_{P_1 \otimes P_2}(f^2) \leq P_2 \left( \text{Ent}_{P_1}(f^2) \right) + P_1 \left( \text{Ent}_{P_2}(f^2) \right). \]

We may now apply the \( \text{LSI}(c_i) \), for \( i \in \{1,2\} \), and the statement of the lemma is straightforward.

\[ \square \]

Lemma 3.16 (Herbst’s argument). Assume that \( P \) satisfies \( \text{LSI}(c) \). Let \( F : \mathbb{R}^n \to \mathbb{R} \) be Lipschitz with constant 1. Then for all \( \lambda \in \mathbb{R} \),

\[ E_P e^{\lambda (F - E_P F)} \leq e^{c \lambda^2}, \]

and so for any \( t \geq 0 \),

\[ P(|F - E_P F| \geq t) \leq 2e^{-\frac{t^2}{e}}. \]
Proof. We use Chernoff bound, for \( \lambda \in \mathbb{R} \),
\[
P(F - E_PF \geq t) \leq e^{-\lambda t}E_pe^{\lambda(F-E_PF)}.
\]
Applied \( \lambda = t/(2c) \), \( F \) and \(-F\), we deduce that the second statement is a consequence of the first statement. We start with the case of \( F \) continuously differentiable. By homogeneity, we can assume \( E_PF = 0 \) and \( \lambda > 0 \). Consider the log-Laplace function
\[
\Lambda(\lambda) = \log E_pe^{2\lambda F}.
\]
Apply the definition of LSI to the function \( f = e^{\lambda F} \). We find
\[
2\lambda E_PF e^{2\lambda F} - E_pe^{2\lambda F} \log \left( E_pe^{2\lambda F} \right) \leq cE_P \sum_{i=1}^{n} \left| \partial_i F e^{\lambda F} \right|^2.
\]
Since
\[
\left( \frac{\Lambda(\lambda)}{\lambda} \right)' = \frac{E_PF e^{2\lambda F}}{\lambda e^{2\lambda F}} - \frac{\log E_pe^{2\lambda F}}{\lambda^2}.
\]
It yields to
\[
\left( \frac{\Lambda(\lambda)}{\lambda} \right)' \leq cL^2
\]
where \( L^2 = \max_{x \in \mathbb{R}^n} \| \nabla F \|^2 = \max_{x \in \mathbb{R}^n} \sum_{i=1}^{n} |\partial_i F(x)|^2 \leq 1 \). Since \( E_PF = 0 \), \( \Lambda(\lambda) = o(\lambda) \) as \( \lambda \downarrow 0 \), we find for all \( \lambda > 0 \),
\[
\Lambda(\lambda) \leq c\lambda^2.
\]
This proves the first statement for \( F \) continuously differentiable. In the general case, refer to \( \text{[AGZ10]} \) or \( \text{[ABC+00]} \) [lemma 7.3.3].

We can now derive powerful concentration inequalities for random matrices with independent entries. From lemma 3.16 and corollary 3.8 we find the following.

**Theorem 3.17** (Concentration of ESD with LSI). Let \( M \in \mathcal{M}_n(\mathbb{C}) \) be an Hermitian random matrix and assume that the variable \((M_{ij})_{1 \leq i, j \leq n}\) satisfies LSI(c) in \( \mathbb{R}^{n^2} \). Then for any \( f : \mathbb{R} \to \mathbb{R} \) such that \( \| f \|_L \leq 1 \) and every \( t \geq 0 \),
\[
\mathbb{P} \left( \left| \int f d\mu_{M/\sqrt{n}} - \mathbb{E} \int f d\mu_{M/\sqrt{n}} \right| \geq t \right) \leq 2 \exp \left( -\frac{n^2t^2}{4c} \right).
\]
As a consequence, we find
\[
\text{Var} \left( n \int f d\mu_{M/\sqrt{n}} \right) = O(\| f \|_L^2).
\]
As we will see, this is the good order of magnitude. The fluctuation of the eigenvalues is thus much smaller than the fluctuation of independent variables.

We can also derive a result for the concentration of single eigenvalues.

**Theorem 3.18** (Concentration of single eigenvalue with LSI). Let \( M \in \mathcal{M}_n(\mathbb{C}) \) be an Hermitian random matrix and assume that the random variable \((M_{ij})_{1 \leq i, j \leq n}\) satisfies LSI(c) in \( \mathbb{R}^{n^2} \). Then for any \( 1 \leq k \leq n \) and \( f : \mathbb{R} \to \mathbb{R} \) such that \( \| f \|_L \leq 1 \) and every \( t \geq 0 \),
\[
\mathbb{P} \left( \left| f(\lambda_k(M/\sqrt{n})) - \mathbb{E} f(\lambda_k(M/\sqrt{n})) \right| \geq t \right) \leq 2 \exp \left( -\frac{nt^2}{4c} \right).
\]
This result is less appealing, it gives the bound
\[
\text{Var} \left( f(\lambda_k(M/\sqrt{n})) \right) = O \left( \frac{1}{n} \right).
\]
However, this time, it is not good enough. For GUE matrices and \( k = \lfloor \alpha n \rfloor \) with \( 0 < \alpha < 1 \), the variance is of order \( O \left( \frac{\log n}{n^2} \right) \).
3.3. CONCENTRATION OF MEASURE FOR RANDOM MATRICES

3.3.3 With Talagrand’s concentration inequality

We start by recalling a corollary of Talagrand’s concentration inequality.

Theorem 3.19 (Talagrand’s concentration inequality). Let $K$ be a convex compact subset of $\mathbb{C} \simeq \mathbb{R}^2$ with diameter $D = \sup_{x,y \in K} |x - y|$. Consider a convex Lipshitz real-valued function $f$ defined on $K^n$. Let $P = P_1 \otimes \cdots \otimes P_n$ be a product measure on $K^n$ and let $M_P(f)$ be the median of $f$ under $P$. Then for any $t > 0$,

$$P(|f - M_P(f)| \geq t) \leq 4 e^{-\frac{t^2}{8D^2 \|f\|_L^2}}.$$  

For a proof in the case $K \subset \mathbb{R}$ see Ledoux [Led01] (with constant 4 instead of 8). If $E_P(f)$ is the mean of $f$ under $P$, with the assumption of the theorem, we find

$$|M_P(f) - E_P(f)| = \int_{-\infty}^{\infty} P(|f - M_P(f)| \geq t) dt \leq 2 \int_{-\infty}^{\infty} e^{-\frac{t^2}{8D^2 \|f\|_L^2}} dt = \sqrt{\frac{2}{\pi}} D \|f\|_L.$$

We may then deduce from Talagrand’s theorem 3.19, an inequality with the $E_P(f)$ in place of $M_P(f)$. In comparison with lemma 3.16, note the extra requirement that $f$ is convex. Hence in order to apply this result to empirical spectral distribution, the following lemma is useful.

Lemma 3.20 (Klein’s lemma). Suppose that $f$ is a real-valued convex function on $\mathbb{R}$. Then the function $F : X \mapsto \text{Tr} f(X)$ on $H_n(\mathbb{C})$ is convex.

Proof. We refer to the proof in [AGZ10], p286.

From theorem 3.19, corollary 3.18 and lemma 3.20 we find the following.

Theorem 3.21 (Concentration of ESD with independent bounded entries). Let $M \in H_n(\mathbb{C})$ be an Hermitian random matrix and assume that the variables $(M_{ij})_{1 \leq i \leq j \leq n}$ are independent variables with support in $K \subset \mathbb{C} \simeq \mathbb{R}^2$ convex with diameter $D$. Then for any convex $f : \mathbb{R} \to \mathbb{R}$, for every $t \geq 0$,

$$\mathbb{P} \left( \left| \int f \, d\mu_{M/\sqrt{n}} - M_P(f) \right| \geq t \right) \leq 4 \exp \left( -\frac{n^2 t^2}{16D^2 \|f\|_L^2} \right),$$

where $M_P(f)$ is the median under $\mathbb{P}$ of $\int f \, d\mu_{M/\sqrt{n}}$.

The restriction that $f$ is convex can be partially lifted if $f$ has a finite number of inflection points.

Corollary 3.22. Let $M \in H_n(\mathbb{C})$ be an Hermitian random matrix and assume that the variables $(M_{ij})_{1 \leq i \leq j \leq n}$ are independent variables with support in $K \subset \mathbb{C} \simeq \mathbb{R}^2$ convex with diameter $D$. Then for any $C^2$ function $f : \mathbb{R} \to \mathbb{R}$ with $k$ inflection points, for every $t \geq 0$,

$$\mathbb{P} \left( \left| \int f \, d\mu_{M/\sqrt{n}} - M_P(f) \right| \geq t \right) \leq 4(k + 1) \exp \left( -\frac{n^2 t^2}{16(k + 1)^2 D^2 \|f\|_L^2} \right),$$

where $M_P(f)$ is the median under $\mathbb{P}$ of $\int f \, d\mu_{M/\sqrt{n}}$. 
Proof. The corollary follows Theorem 3.21 together with two facts:

\[ \{ x_1 + \cdots + x_\ell \geq t \} \subset \bigcup_{1 \leq i \leq \ell} \left\{ x_i \geq \frac{t}{\ell} \right\}; \]

and a function \( f \) with \( k \) inflection points can written as

\[ f = \sum_{i=1}^{k+1} \varepsilon_i f_i \]

where \( \varepsilon_i \in \{-1, 1\} \), \( f_i \) is convex and \( \| f_i \|_L \leq \| f \|_L \). Indeed, this last property can be checked by recursion on \( k \). Let \( x_1 < \cdots < x_k \) be the inflection points of \( f \). Up to considering \( \pm (f(\cdot - x_k) - f(x_k)) \), we may assume without loss of generality that \( x_k = 0 \), \( f(0) = 0 \) and \( f''(x) > 0 \) for \( x > 0 \) (where it exists). We decompose \( f \) as

\[
\begin{align*}
\quad f(x) & = f(x) - f'(0)x + f'(0)x \\
& = \left\{ (f(x) - f'(0)x)1(x < 0) + f'(0)x \right\} + (f(x) - f'(0)x)1(x \geq 0) \\
& = g_1(x) + g_2(x).
\end{align*}
\]

Notice that \( g_1 \) and \( g_2 \) are \( C^1 \) functions. Moreover, \( g_2 \) is convex and, for \( x \geq 0 \), \( g_2'(x) = f'(x) - f'(0) \in [0, f'(x)] \) so that \( \| g_2 \|_L \leq \| f \|_L \). Similarly, \( g_1 \) has \( k-1 \) inflection points and, since \( g_1(x) = f(x) \) on \( (-\infty, 0] \) and \( g_1(x) = f'(0)x \) on \( [0, \infty) \) we find \( \| g_1 \|_L \leq \| f \|_L \). \( \square \)
Chapter 4

A second approach to Wigner Theorem: the resolvent method

In Section 2.3 we have seen that the even moments of Wigner’s semi-circular law are given by the Catalan number. The generating function of the Catalan number satisfies a very simple fixed point equation (2.2.2). This hints that the generating function of moments of ESD of random matrices could be easier to compute than the actual moments. The resolvent method formalizes this ideas.

4.1 Cauchy-Stieltjes transform

4.1.1 Definition and properties

Let \( \mu \) be a finite measure on \( \mathbb{R} \). Define its Cauchy-Stieltjes transform as for all \( z \in \mathbb{C}^+ = \{ z \in \mathbb{C} : \Im(z) > 0 \} \),

\[
g_{\mu}(z) = \int \frac{1}{\lambda - z} d\mu(\lambda).
\]

Note that if \( \mu \) has bounded support we have

\[
g_{\mu}(z) = -\sum_{k \geq 0} z^{-k-1} \int \lambda^k d\mu(\lambda).
\]

The Cauchy-Stieltjes transform is thus essentially the generating function of the moments of the measure \( \mu \).

Lemma 4.1 (Properties of Cauchy-Stieltjes transform). Let \( \mu \) be a finite measure on \( \mathbb{R} \) with mass \( \mu(\mathbb{R}) \leq 1 \).

(i) Analytic: the function \( g_{\mu} \) is an analytic function from \( \mathbb{C}^+ \to \mathbb{C}^+ \).

(ii) Bounded: for any \( z \in \mathbb{C}^+ \), \( |g_{\mu}(z)| \leq (\Im(z))^{-1} \).

The Cauchy-Stieltjes transform characterizes the measure. More precisely, the following holds.

Lemma 4.2 (Inversion of Cauchy-Stieltjes transform). Let \( \mu \) be a finite measure on \( \mathbb{R} \).

(i) For any \( f \in C_0(\mathbb{R}) \),

\[
\int f d\mu = \lim_{t \downarrow 0} \frac{1}{\pi} \int f(x) \Im g_{\mu}(x + it) dx.
\]

(ii) If \( f = 1_I \) with \( I \) is interval and \( \mu(\partial I) = 0 \) the above formula holds.
(iii) For any \( x \in \mathbb{R} \),
\[
\mu(\{x\}) = \lim_{t \downarrow 0} t \Im g_\mu(x + it).
\]

(iv) If \( \mu \) admits a density at \( x \in \mathbb{R} \), then its density is equal to
\[
\lim_{t \downarrow 0} \frac{1}{\pi} \Im g_\mu(x + it).
\]

**Proof.** By linearity, we can assume that \( \mu \) is probability measure. We have the identity
\[
\Im g(x + it) = \int \frac{t}{(\lambda - x)^2 + t^2} d\mu(\lambda).
\]
Hence \( \frac{1}{\pi} \Im g(x + it) \) is the equal to density at \( x \) of the distribution \( (\mu * P_t) \), \( P_t \) is a Cauchy distribution with density
\[
P_t(x) = \frac{t}{\pi(x^2 + t^2)}.
\]
In other words,
\[
\frac{1}{\pi} \int f(x) \Im g_\mu(x + it) dx = Ef(X + tY),
\]
where \( X \) has law \( \mu \) and is independent of \( Y \) with distribution \( P_t \). The statements follow easily. \( \square \)

### 4.1.2 Cauchy-Stieltjes transform and weak convergence

The convergence of Cauchy-Stieltjes transform is equivalent to the weak convergence.

**Corollary 4.3.** Let \( \mu \) and \( (\mu_n)_{n \geq 1} \) be a sequence of real probability measures. The following are equivalent

(i) As \( n \to \infty \), \( \mu_n \rightharpoonup \mu \).

(ii) For all \( z \in \mathbb{C}_+ \), as \( n \to \infty \), \( g_{\mu_n}(z) \to g_\mu(z) \).

(iii) There exists a set \( D \subset \mathbb{C}_+ \) with an accumulation point such that for all \( z \in D \), as \( n \to \infty \), \( g_{\mu_n}(z) \to g_\mu(z) \).

**Proof.** Statement "(i) implies (ii)" follows from the definition of weak convergence applied to the real and imaginary part of \( f(\lambda) = (\lambda - z)^{-1} \). Statement "(ii) implies (iii)" is trivial. For statement "(iii) implies (i)", from Helly selection theorem, the sequence \( (\mu_n) \) is relatively compact for the vague convergence. Let \( \nu \) be such vague limit, it is a finite measure with mass at most 1. By assumption, for any \( z \in D \), \( g_\nu(z) = g_\mu(z) \). Two analytic functions equal on a set with an accumulation point are equal on their domain (principle of analytic extension). Hence, by lemma \[4.1\] for any \( z \in \mathbb{C}_+ \), \( g_\mu(z) = g_\nu(z) \). By lemma \[4.2\] we deduce that \( \nu \) is a probability measure and \( \nu = \mu \). \( \square \)

### 4.1.3 Cauchy-Stieltjes transform of the semi-circular law

The Cauchy-Stieltjes semi-circular distribution \( \mu_{sc} \) satisfies the fixed point for all \( z \in \mathbb{C}_+ \),
\[
g_{\mu_{sc}}(z) = \frac{1}{z + g_\mu(z)} \quad \text{or} \quad g_{\mu_{sc}}(z)^2 + zg_{\mu_{sc}}(z) + 1 = 0. \tag{4.1.1}
\]
Let \( z \mapsto \sqrt{z} \) be the analytical continuation of \( x \mapsto \sqrt{x} \) on \( \mathbb{C} \setminus \mathbb{R}_- \) with a positive imaginary part. We find
\[
g_{\mu_{sc}}(z) = \frac{-1 + \sqrt{1 - 4z}}{2}.
\]
4.2 Resolvent

4.2.1 Spectral measure at a vector

Let $A \in \mathcal{M}_n(\mathbb{C})$ and $\phi \in \mathbb{C}^n$ be a vector with unit $\ell_2$-norm, $\|\phi\|_2 = 1$.

The spectral theorem guarantees the existence of $(v_1, \cdots, v_n)$, an orthonormal basis of $\mathbb{C}^n$ of eigenvectors of $A$, that is, for any $1 \leq i \leq n$, $Av_i = \lambda_i(A)v_i$. The spectral measure with vector $\phi$ is the real probability measure defined by

$$
\mu_A^\phi = \sum_{k=1}^n |\langle v_k, \phi \rangle|^2 \delta_{\lambda_k(A)}.
$$

(4.2.1)

It may also be defined as the unique probability measure $\mu_A^\phi$ such that

$$
\int \lambda^k d\mu_A^\phi(\lambda) = \langle \phi, A^k \phi \rangle \quad \text{for any integer } k \geq 1.
$$

(4.2.2)

If $\phi = (\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}})$, we have the identity $\mu_A = \mu_A^\phi$. Also, if $(e_1, \cdots, e_n)$ is the canonical basis of $\mathbb{C}^n$, summing (4.2.1), we find, with $\phi_i = \langle \phi, e_i \rangle$,

$$
\mu_A^\phi = \sum_{i=1}^n |\phi_i|^2 \mu_A^{e_i} \quad \text{and} \quad \mu_A = \frac{1}{n} \sum_{i=1}^n \mu_A^{e_i}.
$$

(4.2.3)

4.2.2 Resolvent matrix

If $A \in \mathcal{H}_n(\mathbb{C})$ and $z \in \mathbb{C}_+ = \{z \in \mathbb{C} : \Im(z) > 0\}$, then $A - zI$ is invertible. We define the resolvent of $A$ as the function $R : \mathbb{C}_+ \rightarrow \mathcal{M}_n(\mathbb{C})$,

$$
R(z) = (A - zI)^{-1}.
$$

We have the identity

$$
\langle \phi, R(z) \phi \rangle = \int \frac{1}{\lambda - z} d\mu_A^\phi(\lambda) = g_{\mu_A^\phi}(z),
$$

(4.2.4)

where $\mu_A^\phi$ is the spectral measure with vector $\phi$. Also,

$$
g_{\mu_A}(z) = \frac{1}{n} \Tr(R(z)).
$$

Lemma 4.4 (Properties of the resolvent matrix). Let $A \in \mathcal{H}_n(\mathbb{C})$ and $R(z) = (A - zI)^{-1}$ be its resolvent. For any $z \in \mathbb{C}_+$, $1 \leq i, j \leq n$,

(i) Analytic : $z \mapsto R(z)_{ij}$ is an analytic function on $\mathbb{C}_+ \rightarrow \mathbb{C}$.

(ii) Bounded : $\|R(z)\|_{2\rightarrow 2} \leq \Im(z)^{-1}$.

(iii) Normal : $R(z)^*R(z) = R(z)R(z)^*$.

Proof. All properties come from the decomposition

$$
R(z) = \sum_{k=1}^n \frac{1}{\lambda_k(A) - z} v_kv_k^*,
$$

where $v_k$ is an orthogonal basis of eigenvectors of $A$. \hfill \square
4.2.3 Perturbation inequalities of resolvent

Rank inequalities

**Lemma 4.5** (Rank inequality for resolvent). Let $A$, $B$ in $\mathcal{H}_n(\mathbb{C})$. Then, if $z \in \mathbb{C}^+$, $R_A(z) = (A - zI_n)^{-1}$ and $R_B(z) = (B - zI_n)^{-1}$,
\[
\sum_{k=1}^{n} |R_A(z)_{kk} - R_B(z)_{kk}| \leq \frac{2\text{rank}(A - B)}{2\text{Im} z}.
\]

*Proof.*

There are variants of the above interlacing inequality. The following can be proved as the above lemma.

**Corollary 4.6** (Rank inequality for resolvent of minor). Let $A \in \mathcal{H}_n$ and $B$ be a minor of $A$ where a subset $\Lambda$ of rows and column have been removed. Then, if $z \in \mathbb{C}^+$, $R_A(z) = (A - zI_n)^{-1}$ and $R_B(z) = (B - zI_n)^{-1}$,
\[
\sum_{k \in \{1, \cdots, n\} \setminus \Lambda} |R_A(z)_{kk} - R_B(z)_{kk}| \leq \frac{2|\Lambda|}{2\text{Im} z}.
\]

($B$ is seen as a linear map on $H = \text{span}(e_k : k \in \{1, \cdots, n\} \setminus \Lambda)$).

*Proof.* Consider the matrix $B' \in \mathcal{M}_n(\mathbb{C})$ obtained from $B$, by setting $B'e_k = 0$ for $k \in \Lambda$ and $B'e_k = Be_k$ otherwise. We obviously have $(B - zI_n)^{-1}_{kk} = (B' - zI_n)^{-1}_{kk}$ for $k \notin \Lambda$. Moreover by construction $\text{rank}(B' - A) \leq |\Lambda|$. We may then apply lemma 4.5.

Norm inequalities

The reader may not be surprised that in the above arguments we have used rank inequalities. Using Hoffman-Wielandt inequality, similar inequalities can be derived in term of norm inequalities.

4.2.4 Resolvent complement formula

The Schur complement is simply a block inversion by part of an invertible matrix.

**Lemma 4.7** (Schur’s complement formula). Let $A \in \mathcal{M}_n(\mathbb{C})$ be an invertible matrix. Set
\[
A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad A^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}
\]
where $A_{11}, B_{11} \in \mathcal{M}_p(\mathbb{C})$. Then, if $A_{22}$ and $B_{11}$ are invertible, we have
\[
B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}.
\]

**Corollary 4.8** (Resolvent complement formula). Let $n \geq 2$, $A = (a_{ij})_{1 \leq i, j \leq n} \in \mathcal{H}_n(\mathbb{C})$, $z \in \mathbb{C}^+$ and $R = (A - zI_n)^{-1}$. For any $1 \leq i \leq n$,
\[
R_{ii} = -(z - a_{ii} + (a_i, R_i a_i))^{-1},
\]
where $a_i = (a_{ij})_{j \neq i} \in \mathbb{C}^{n-1}$, $R_i = (A_i - zI_{n-1})^{-1}$ and $A_i \in \mathcal{H}_{n-1}(\mathbb{C})$ is the principal minor of $A$ where the $i$-th row and column have been removed.
4.3 Resolvent method for random matrices

In this section, we will present an exemple the resolvent method for random matrix. This method will be based on two components: a probabilistic component, the concentration of bilinear forms and a linear algebra component, the Schur complement formula.

4.3.1 Concentration for bilinear forms

**Lemma 4.9** (Variance of Bilinear form of independent vectors). Let $A \in M_n(\mathbb{C})$ and $X = (X_1, \ldots, X_n) \in \mathbb{C}^n$ be a vector of centered and independent variables with $\mathbb{E}|X_i|^2 = 1$ and $\text{Var}(|X_i|^2) \leq \kappa$ for $1 \leq i \leq n$. Then

$$\mathbb{E}\langle X, AX \rangle = \text{Tr}A \quad \text{and} \quad \text{Var}(\langle X, AX \rangle) \leq 2\text{Tr}AA^* + \kappa n \sum_{i=1}^{n} |A_{ii}|^2.$$

*Proof.* We have

$$\langle X, AX \rangle = \sum_{1 \leq i, j \leq n} \bar{X}_i A_{ij} X_j.$$

This yields to

$$\mathbb{E}\langle X, AX \rangle = \sum_{i=1}^{n} \mathbb{E}|X_i|^2 A_{ii} = \text{Tr}A$$

and

$$\text{Var}(\langle X, AX \rangle) = \sum_{i_1, j_1, i_2, j_2} A_{i_1 j_1} A_{i_2 j_2} \mathbb{E}X_{i_1} X_{j_2} X_{i_2} X_{j_2} - \sum_{i,j} \mathbb{E}|X_i|^2 \mathbb{E}|X_j|^2 A_{ii} A_{jj}.$$

The first sum is non zero only if $(i_1, i_2) = (j_1, j_2)$, $(i_1, j_1) = (i_2, j_2)$ or $(i_1, j_1) = (j_2, i_2)$. We get

$$\text{Var}(\langle X, AX \rangle) \leq \kappa \sum_{i=1}^{n} |A_{ii}|^2 + \sum_{i \neq j} |A_{ij}|^2 + \sum_{1 \leq i, j \leq n} |A_{ij}|^{1/2} |A_{ji}|^{1/2} \mathbb{E}X_i^2 \mathbb{E}X_j^2.$$

The second term is equal $\text{Tr}(AA^*)$, while the third term is upper bounded by $\text{Tr}(AA^*)$ from Cauchy-Schwarz inequality.

With more moments assumption, it is of course possible to strengthen lemma 4.9. For example, for entries with sub-gaussian tail, this is topic of the Hanson-Wright theorem.

4.4 Limit empirical spectral distribution of band matrices

In this section, we show the resolvent method at work with the band matrices. For each integer $n \geq 1$, we assume that $(X_{ij})_{1 \leq i \leq j}$ are independent real centered variables with variance 1. We set, for $1 \leq i \leq j \leq n$, $Y_{ji} = Y_{ij}$ and

$$Y_{ij} = X_{ij} \int_{Q_{ij}} \frac{\sigma(x,y)}{|Q_{ij}|} dxdy,$$

where $Q_{ij} = [(i-1)/n, i/n] \times [(j-1)/n, j/n]$ and $\sigma : [0,1]^2 \rightarrow [0,1]$ is a measurable function such that

$$\sigma(x,y) = \sigma(y,x).$$

We consider the Hermitian matrix

$$Y_n = (Y_{ij})_{1 \leq i, j \leq n}.$$

We assume that all these matrices are defined on a common probability space.
Theorem 4.10 (ESD of matrices with variance profile). There exists a probability measure $\mu_\sigma$ depending on $\sigma$ such that a.s.

$$\mu_{Y/\sqrt{n}} \sim \mu_\sigma.$$ 

The Cauchy-Stieltjes transform $g_{\mu_\sigma}$ of $\mu$ is given by the formula

$$g_{\mu_\sigma}(z) = \int_0^1 g(x, z)dx,$$

where the $[0, 1] \times \mathbb{C}_+ \rightarrow \mathbb{C}_+$ map, $g : (x, z) \mapsto g(x, z)$ satisfies: for a.a. $x \in [0, 1]$, $z \mapsto g(x, z)$ analytic on $\mathbb{C}_+$ and for each $z \in \mathbb{C}_+$ with $\Im(z) > 1$, $x \mapsto g(x, z)$ is the unique function in $L^1([0, 1]; \mathbb{C}_+)$ solution of the equation, for a.a. $x \in [0, 1]$,

$$g(x, z) = -\left(z + \int_0^1 \sigma^2(x, y)g(y, z)dy\right)^{-1}. \quad (4.4.1)$$

Note that by analyticity, for $0 < \Im(z) \leq 1$, $x \mapsto g(x, z)$ is also a solution of (4.4.1) (which may however a priori not be unique in $L^1([0, 1]; \mathbb{C}_+)$). A typical application of theorem 4.10 is the following.

Corollary 4.11 (ESD of band matrices). Assume that for a.a. $x \in [0, 1]$, $\int_0^1 \sigma^2(x, y)dx = b$ for some $b \in (0, 1]$. Then

$$\mu_{Y/\sqrt{bn}} \sim \mu_{sc}.$$

The above corollary follows from noticing that the semi-circular law satisfies the fixed point equation (4.1.1) and the unicity statement in theorem 4.10. Using Hoffman-Wielandt inequality, it may be used to prove that a.s.

$$\mu_{Z/\sqrt{bn}} \sim \mu_{sc},$$

where $Z_{ij} = X_{ij}1(\max(|i-j|, |n-i-j|) \leq bn)$. In fact, from the proof of theorem 4.10, it can be seen that this last result holds for $b = b(n) \rightarrow 0$, as soon as $b(n)n \rightarrow \infty$ (see exercice 4.13).

Let $1 \leq p \leq n$ and define the matrix in $\mathcal{M}_{p,n}(\mathbb{R})$,

$$X_n = (X_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}.$$ 

Corollary 4.12 (Marcenko-Pastur law). Assume $p(n)/n \rightarrow c \in (0, 1]$ then a.s

$$\mu_{XX^*/n} \sim \mu_{MP}.$$ 

where, $b_- = (1 - \sqrt{c})^2$, $b_+ = (1 + \sqrt{c})^2$ and

$$\mu_{MP}(dx) = \frac{1}{2\pi x}\sqrt{(x - b_-)(b_+ - x)}1_{b_- \leq x \leq b_+}dx$$

Note that from the above corollary, it is also possible to deal with case $c > 1$. It suffices to reverse the role of $p$ and $n$ and notice that, for $1 \leq p \leq n$

$$n\mu_{XX^*} = p\mu_{XX^*} + (n - p)\delta_0.$$ 

Proof of corollary 4.12. Consider the block matrix in $\mathcal{H}_{n+p}(\mathbb{C})$,

$$Z = \begin{pmatrix} 0_n & X^* \\ X & 0_p \end{pmatrix}. \quad (4.4.2)$$

If $0 \leq \lambda_1 \leq \cdots \leq \lambda_p$ are the eigenvalues of $XX^*$ with $\lambda_1 = \cdots = \lambda_m = 0$, $\lambda_{m+1} > 0$, then the non-zero eigenvalues of $Z$ are

$$\pm \sqrt{\lambda_{m+1}}, \ldots, \pm \sqrt{\lambda_p}.$$
4.4. LIMIT EMPIRICAL SPECTRAL DISTRIBUTION OF BAND MATRICES

In particular, if \( \mu_{Z/\sqrt{n+p}} \) converges weakly toward a limit measure \( \nu \) with

\[
\nu = \frac{1 - c}{1 + c} \delta_0 + \frac{2c}{1 + c} \hat{\nu},
\]

where \( \hat{\nu} \) is a symmetric probability measure on \( \mathbb{R} \) with density \( f \), then \( \mu_{XX^*/n} \) converges weakly to \( \mu \) with density on \((0, \infty)\) given by

\[
d\mu(x) = f\left(\sqrt{x}\right)\sqrt{x} \, dx.
\]

Now, coming back to (4.4.2), we introduce the \([0, 1/2] \to [0, 1]\) function

\[
\sigma(x, y) = 1 \left( 0 < x < \frac{1}{1 + c} \right) 1 \left( \frac{1}{1 + c} < y < 1 \right) + 1 \left( \frac{1}{1 + c} < x < 1 \right) 1 \left( 0 < y < \frac{1}{1 + c} \right).
\]

Note that \( \sigma(x, y) = \sigma(y, x) \) and we may consider the associated matrix \( Y \) as in theorem 4.10.

It is an exercise to compute explicitly \( \mu_{\sigma} \) in this case. We may then use Hoffman-Wielandt inequality lemma 3.6 to check that \( \mu_{Z/\sqrt{n+p}} \) and \( \mu_{Y/\sqrt{n+p}} \) are asymptotically close.

4.4.1 Proof of theorem 4.10

The proof of theorem 4.10 is a typical instance of the resolvent method. In the first step of the proof, we check tightness of \( \mu_{Y/\sqrt{n}} \) and that for each \( z \in \mathbb{C}_+ \), a.s.

\[
\frac{1}{n} \sum_{i=1}^{n} R_{ii}(z) \to \int_{0}^{1} g(x, z) \, dx,
\]

where, for each \( z \in \mathbb{C}_+ \), \( g : x \mapsto g(x, z) \) is a fixed point of the \( L^1([0, 1]; \mathbb{C}_+) \to L^1([0, 1]; \mathbb{C}_+) \) map

\[
F_{z, \sigma}(g)(x) = -\left( z + \int \sigma^2(x, y) g(y) \, dy \right)^{-1}.
\]

Observe that (4.4.3) implies that, if \( D \) is countable dense set in \( \mathbb{C}_+ \), a.s. for all \( z \in D \), (4.4.3) holds. Then, in a second step, we prove the uniqueness of the solution of \( F_{z, \sigma}(g) = g \) for \( \Im(z) \) large enough. Now, consider a converging subsequence of \( (\mu_{Y/\sqrt{n}})_{n \geq 1} \) to \( \mu \). Invoking (4.2.4) and corollary 4.3 we conclude that a.s. for \( z \in \mathbb{C}_+ \), \( g_{\mu}(z) = \int_{0}^{1} g(x, z) \, dx \). By the unicity of the limit and a new application of corollary 4.3 it will conclude the proof of theorem 4.10. In the sequel, the parameter \( z \in \mathbb{C}_+ \) is fixed and, for ease of notation, we will often omit it.

Tightness

We write

\[
\int \lambda^2 d\mu_{Y/\sqrt{n}} = \frac{1}{n^2} \text{Tr} YY^* \leq \frac{1}{n^2} \sum_{i,j} |X_{ij}|^2.
\]

From the law of large numbers, we deduce that a.s.

\[
\lim_{n \to \infty} \sup \int \lambda^2 d\mu_{Y/\sqrt{n}} \leq 1.
\]

Hence, a.s., the sequence of probability measures \( (\mu_{Y/\sqrt{n}})_{n \geq 1} \) is tight.
Concentration and truncation

Arguing as in proposition 3.9, it is sufficient to prove theorem 4.10 when \( X_{ij} \) are bounded by say \( K \). Moreover, by concentration lemma 3.17, it is sufficient to prove that

\[
E \frac{1}{n} \sum_{i=1}^{n} R_{ii} \rightarrow \int_{0}^{1} g(x, z) dx,
\]

and (4.4.3) will follow.

Approximation of \( \sigma \)

Let \((P_{k\ell})_{1 \leq k, \ell \leq L}\) be the usual partition of \([0, 1]^2\) into squares of size \( 1/L^2 \). Define

\[
\rho = \sum_{1 \leq k, \ell \leq L} \rho_{k\ell} 1_{P_{k\ell}},
\]

where \( \rho_{k\ell} = L^2 \int_{P_{k\ell}} \sigma(x, y) dy dx \). We define the matrix \( Z \) by

\[
Z_{ij} = X_{ij} \sum_{1 \leq k, \ell \leq L} \rho_{k\ell} 1_{P_{k\ell}} \left( \frac{i}{n}, \frac{j}{n} \right), \tag{4.4.4}
\]

We have

\[
\frac{1}{n^2} \text{Tr}(Z-Y)^2 \leq K^2 \int_{[0,1]^2} |\sigma^2(x,y)-\rho^2(x,y)| dy dx + K^2 \int_{[0,1]^2} |\sigma^2(x,y)^2-\dot{\sigma}^2(x,y)| dy dx + O \left( \frac{1}{n^2} \right),
\]

where

\[
\dot{\sigma} = \sum_{1 \leq i,j \leq n} \sigma_{ij} 1_{Q_{ij}},
\]

and \( \sigma_{ij} = n^2 \int_{Q_{ij}} \sigma(x,y) dy dx \). From Lebesgue theorem, for a.a \((x,y)\), \( \sigma(x,y) - \dot{\sigma}(x,y) \to 0 \) and as \( L \to \infty \), \( \sigma(x,y) - \rho(x,y) \to 0 \). Hence, by dominated convergence we deduce that

\[
\|\sigma^2 - \dot{\sigma}^2\|_1 \to_{n \to \infty} 0 \quad \text{and} \quad \|\sigma^2 - \rho^2\|_1 \to_{L \to \infty} 0
\]

In particular,

\[
\limsup_{n \to \infty} \frac{1}{n^2} \text{Tr}(Z-Y)^2 \leq \varepsilon(L),
\]

for some function \( \varepsilon \) going to 0 as \( L \) goes to infinity.

Approximate fixed point equation

Consider the matrix \( Z \) given by (4.4.4). The objective is to prove that the resolvent of \( Z \) satisfies nearly a fixed point equation. To this end, we use Schur complement formula, corollary 4.8

\[
R_{ii} = - \left( z - \frac{Z_{ii}}{\sqrt{n}} + \frac{1}{n} (Z_i, R^{(i)} Z_i) \right)^{-1},
\]

where \( Z_i = (Z_{ij})_{j \neq i} \) and \( R^{(i)} = (Z^{(i)} - zI)^{-1} \) is the resolvent of the minor matrix \( Z^{(i)} \) obtained from \( Z \) where the \( i \)-th row and column have been removed.

Notice that if \( z, w, w' \in \mathbb{C}_+ \), then

\[
\left| \frac{1}{z+w} - \frac{1}{z+w'} \right| \leq \frac{|w-w'|}{2m(z)^2}. \tag{4.4.5}
\]
Since \( \langle Z_i, R^{(i)} Z_i \rangle \in \mathbb{C}_+ \), \( R^{(i)}_{jj} \in \mathbb{C}_+ \) and \( \mathbb{E}|Z_{ij}|^2 = \rho \left( \frac{i}{n}, \frac{j}{n} \right) \), we find

\[
\left| R_{ii} + \left( z + \frac{1}{n} \sum_{j \neq i} \rho \left( \frac{i}{n}, \frac{j}{n} \right) R^{(i)}_{jj} \right)^{-1} \leq \frac{1}{3m(z)^2} \left( \frac{|Z_{ii}|}{\sqrt{n}} + \frac{1}{n} \right) \left| \frac{\langle Z_i, R^{(i)} Z_i \rangle - \sum_{j \neq i} (\mathbb{E}|Z_{ij}|^2) R^{(i)}_{jj} }{z} \right| .
\]

Now, by construction, the vector \((Z_{ij})_j\) is independent of \(R^{(i)}\). We condition on \(R^{(i)}\) and use lemma 4.9, we deduce that in \(L^2(\mathbb{P})\),

\[
R_{ii} + \left( z + \frac{1}{n} \sum_{j \neq i} \rho \left( \frac{i}{n}, \frac{j}{n} \right) R^{(i)}_{jj} \right)^{-1} \to 0.
\]

We define \(N_k = \{1 \leq i \leq n : (k-1)/L < i/n \leq k/L\}\) and \(N^{(i)}_k = N_k \setminus \{i\}\). So that \(|N_k|/n \to 1/L\). If \(i \in N_k\), it yields to, in \(L^2(\mathbb{P})\),

\[
\left( z + \frac{1}{L} \sum_{\ell=1}^L \rho^2 \left( \frac{\ell}{L} \right) G^{(i)}_{\ell} \right)^{-1} \to 0,
\]

where we have defined

\[
G^{(i)}_k = \frac{1}{|N_k|} \sum_{j \in N^{(i)}_k} R^{(i)}_{jj}.
\]

Now, from finite rank inequality

\[
|G_k - G^{(i)}_k| \leq \frac{2}{3m(z)|N_k|},
\]

and the proof of concentration lemma 3.17 gives that

\[
\mathbb{E}|G_k - \mathbb{E}G_k|^2 = O \left( \frac{n}{3m(z) |N_k|^2} \right) = O \left( \frac{L^2}{3m(z)^2 n} \right).
\]

It conclusion, using again (4.4.5), for any \(1 \leq k \leq L\),

\[
\mathbb{E}G_k + \left( z + \frac{1}{L} \sum_{\ell=1}^L \rho^2 \left( \frac{\ell}{L} \right) \mathbb{E}G_{\ell} \right)^{-1} \to 0,
\]

and

\[
\mathbb{E} \frac{1}{L} \sum_{k=1}^L G_k = \mathbb{E} \frac{1}{n} \sum_{i=1}^n R_{ii}.
\]

**Unicity of fixed point equation**

Consider the function \(\bar{G} : [0, 1] \to \mathbb{C}_+\) given

\[
\bar{G}(x) = \frac{1}{L} \sum_{k=1}^L 1_{\frac{k-1}{L} < x \leq \frac{k}{L}} \mathbb{E}G_k.
\]

Consider an accumulation point of the vector \((\mathbb{E}G_1, \cdots, \mathbb{E}G_L)\), say \((g_1, \cdots, g_L)\). Then \(\bar{G}\) converges in \(L^\infty\)-norm to

\[
g_\rho(x) = \frac{1}{L} \sum_{k=1}^L 1_{\frac{k-1}{L} < x \leq \frac{k}{L}} g_k.
\]
By [4.4.6], \( g_\rho \) satisfies the fixed point equation, for all \( x \in [0, 1] \),
\[
g = F_{z,\rho}(g),
\]
with
\[
F_{z,\rho}(g)(x) = - \left( z + \int \rho(x, y)^2 g(y) dy \right)^{-1}.
\]

If \( g, h \in L^1([0, 1]; \mathbb{C}_+) \), we find,
\[
|F_{z,\rho}(g)(x) - F_{z,\rho}(h)(x)| \leq \frac{\int \rho^2(x, y)|g(y) - h(y)|dy}{\Im(z)^2} \leq \frac{\|g - h\|_1}{\Im(z)^2},
\]
where we have used again [4.4.5] and \( \rho(x, y) \leq 1 \). In particular,
\[
\|F_{z,\rho}(g) - F_{z,\rho}(h)\|_1 \leq \frac{\|g - h\|_1}{\Im(z)^2}.
\]

Hence for \( \Im(z) > 1 \), \( F_{z,\rho} \) is a contraction on the Banach space \( L^1([0, 1]; \mathbb{C}_+) \). Hence there is a unique solution of the fixed point
\[
g = F_{z,\rho}(g).
\]

The same argument works for the functions \( \sigma \) and its associated map \( F_{z,\rho} \). Now similarly, we have, if \( g \in L^\infty([0, 1]; \mathbb{C}_+) \),
\[
\|F_{z,\sigma}(g) - F_{z,\rho}(g)\|_1 \leq \frac{\int |\rho^2(x, y) - \sigma^2(x, y)||g(y)|dy}{\Im(z)^2} \leq \frac{\|g\|_\infty\|\sigma^2 - \rho^2\|_1}{\Im(z)^2},
\]
In particular, if \( g_\sigma \) is the unique fixed point \( g = F_{z,\sigma}(g) \), since \( \|g_\rho\|_\infty \leq 1/\Im(z) \), we deduce
\[
\|g_\sigma - g_\rho\|_1 = \|F_{z,\sigma}(g_\sigma) - F_{z,\rho}(g_\rho)\|_1 \leq \|F_{z,\sigma}(g_\sigma) - F_{z,\sigma}(g_\rho)\|_1 + \|F_{z,\sigma}(g_\rho) - F_{z,\rho}g_\rho\|_1
\leq \|g_\sigma - g_\rho\|_1 + \frac{\|\sigma^2 - \rho^2\|_1}{\Im(z)^3},
\]
This gives
\[
\|g_\sigma - g_\rho\|_1 \leq \frac{\|\sigma^2 - \rho^2\|_1}{\Im(z)^2(\Im(z) - 1)},
\]
As already pointed, \( \|\rho^2 - \sigma^2\|_1 \to 0 \) as \( L \to \infty \).

**End of proof**

In summary, we have proved the following, fix \( \varepsilon > 0 \), \( z \in \mathbb{C}_+ \) with \( \Im(z) > 1 \). We have, a.s., for all \( n \) large enough,
\[
|g_{\mu_{\gamma,\sqrt{\pi}}}(z) - E g_{\mu_{\gamma,\sqrt{\pi}}}(z)| \leq \varepsilon.
\]

We may fix \( L \) large enough so that
\[
|g_{\mu_\rho}(z) - g_{\mu_\sigma}(z)| \leq \varepsilon.
\]
Then, for all \( n \) large enough,
\[
|E g_{\mu_{\sqrt{\pi}}}(z) - E g_{\mu_{\gamma,\sqrt{\pi}}}(z)| \leq \varepsilon,
\]
and
\[
|E g_{\mu_{\sqrt{\gamma}}}(z) - g_{\mu_\rho}(z)| \leq \varepsilon.
\]
This concludes the proof of theorem [4.10]
4.4. LIMIT EMPIRICAL SPECTRAL DISTRIBUTION OF BAND MATRICES

Exercise 4.13 (Band matrices). Consider the Hermitian matrix $Z = (Z_{ij})_{1 \leq i,j \leq n}$ with $Z_{ij} = X_{ij} 1(|i - j| \leq bn)$, with $b(n) \to 0$ and $nb(n) \to \infty$. Using the resolvent method, prove that a.s. $\mu_{Z/\sqrt{nb}} \Rightarrow \mu_{sc}$.

Exercise 4.14 (Adjacency matrix of Erdős-Rényi graphs). Consider the adjacency matrix $A$ of $G(n, c/n)$ with $0 \leq c(n) \leq n$ and $c(n)n \to \infty$. Namely $(A_{ij})_{1 \leq i < j \leq n}$ are i.i.d. $\{0, 1\}$ Bernoulli random variables with mean $c/n$. Using the resolvent method, prove that a.s. $\mu_{A/\sqrt{nc}} \Rightarrow \mu_{sc}$. 
Chapter 5

Resolvent method with matrix differentiation formulas

5.1 Matrix differentiation formulas

We identify $\mathcal{H}_n(\mathbb{C})$ with $\mathbb{R}^{n^2}$. Then, if $\Phi : \mathcal{H}_n(\mathbb{C}) \mapsto \mathbb{C}$ is a continuously differentiable function, we define $\partial_{jk}\Phi(X)$ as the derivative with respect to $\Re(X_{jk})$, and for $1 \leq j \neq k \leq n$, $\partial'_{jk}\Phi(X)$ as the derivative with respect to $\Im(X_{jk})$.

Define the resolvent $R_A = (A - z)^{-1}$, $z \in \mathbb{C}^+$. From the resolvent formula

$$R_A - R_B = R_A(B - A)R_B,$$

valid for any matrices $A, B \in \mathcal{H}_n(\mathbb{C})$. A simple computation shows that if $1 \leq j, k \leq n$, and $1 \leq a \neq b \leq n$, then

$$\partial_{ab}R_{jk} = -(R_{ja}R_{bk} + R_{jb}R_{ak}) \quad \text{and} \quad \partial'_{ab}R_{jk} = -i(R_{ja}R_{bk} - R_{jb}R_{ak}),$$

while if $1 \leq a \leq n$, then

$$\partial_{aa}R_{jk} = -R_{ja}R_{ak}.$$ 

5.2 Gaussian differentiation formulas

We consider a Wigner matrix $X = (X_{ij})_{1 \leq i,j \leq n}$. We assume that

(A1) $(\Re(X_{12}), \Im(X_{12}))$ is a centered Gaussian vector in $\mathbb{R}^2$ with covariance $K \in \mathcal{H}_2(\mathbb{R})$, $\text{Tr}(K) = 1$.

(A2) $X_{11}$ is a centered Gaussian in $\mathbb{R}$ with variance $\sigma^2$.

We recall the Gaussian integration by part formula.

Lemma 5.1. Let $G$ be a centered Gaussian vector in $\mathbb{R}^p$ with covariance matrix $\Sigma = \mathbb{E}GG^*$. For any continuously differentiable function $F : \mathbb{R}^p \mapsto \mathbb{R}$, with $\mathbb{E}\|\text{grad}F(G)\|_2 < \infty$,

$$\mathbb{E}F(G)G = \Sigma \mathbb{E}\text{grad}F(G).$$

(5.2.1)

The use of Gaussian integration by part in random matrix theory was initiated by Khorunzhy, Khoruzhenko and Pastur [KKP96].

Set $M = X/\sqrt{n}$, so that

$$R = (X/\sqrt{n} - z)^{-1}.$$
CHAPTER 5. RESOLVENT DIFFERENTIATION

Using (5.2.1) we get, for \(0 \leq a \neq b \leq n\), and all \(j,k\):

\[
\mathbb{E} R_{jk} X_{ab} = \frac{1}{\sqrt{n}} \mathbb{E} \left[ K_{11} \partial_{ab} R_{jk} + K_{12} \partial'_{ab} R_{jk} + \frac{i K_{21} \partial_{ab} R_{jk}}{\sqrt{n}} + \frac{i K_{22} \partial'_{ab} R_{jk}}{\sqrt{n}} \right]
\]

\[
= -\frac{1}{\sqrt{n}} \mathbb{E} \left[ (K_{11} - K_{22} + i K_{12} + i K_{21}) R_{ja} R_{bk} + (K_{11} + K_{22} - i K_{12} + i K_{21}) R_{jb} R_{ak} \right]
\]

\[
= -\frac{1}{\sqrt{n}} \mathbb{E} \left( \gamma R_{ja} R_{bk} + R_{jb} R_{ak} \right), \quad (5.2.2)
\]

where at the last line, we have used the symmetry of \(K\) and \(\text{Tr}(K) = 1\), together with the notation

\[
\gamma = K_{11} - K_{22} + 2i K_{12} = \mathbb{E} X_{12}^2.
\]

Notice that \(|\gamma| \leq 1\). Similarly, for \(a = b\) one has

\[
\mathbb{E} R_{jk} X_{aa} = -\frac{\sigma^2}{\sqrt{n}} \mathbb{E} R_{ja} R_{ak}. \quad (5.2.3)
\]

For further use, we also set

\[
\kappa = \sigma^2 - 1 - \gamma.
\]

In the GUE and GOE case \(\kappa = 0\) while \(\gamma\) is equal respectively to 0 and 1.

Notice that in this case the dependency on \(z\) is explicit in our notation.

\[
-\frac{1}{n} \mathbb{E} R = R \left( \frac{X}{\sqrt{n}} - zI - \frac{X}{\sqrt{n}} \right) = I - \frac{1}{\sqrt{n}} RX.
\]

Hence, for \(1 \leq j, k \leq n\), using (5.2.2)-(5.2.3),

\[
-\frac{1}{n} \mathbb{E} R_{jk} = \delta_{jk} - \frac{1}{\sqrt{n}} \sum_{1 \leq a \leq n} \mathbb{E} [R_{ja} X_{ak}]
\]

\[
= \delta_{jk} - \frac{1}{n} \sum_{1 \leq a \leq n} \mathbb{E} [R_{ja} R_{aa}] + \frac{\gamma}{n} \sum_{1 \leq a \neq k \leq n} \mathbb{E} [R_{ja} R_{ka}] + \frac{(\sigma^2 - 1)}{n} \mathbb{E} [R_{jk} R_{kk}].
\]

We set

\[
g = g_{\mu_X / \sqrt{n}}(z) = \frac{1}{n} \text{Tr}(R), \quad \bar{g} = \mathbb{E} g, \quad \underline{g} = g - \mathbb{E} g,
\]

and consider the diagonal matrix \(D\) with \(D_{jk} = 1_{j=k} R_{jk}\). We find

\[
-\frac{1}{n} \mathbb{E} R = I + \mathbb{E} [gR] + \frac{1}{n} \mathbb{E} [R(\kappa D + \gamma R^\top)].
\]

Subtracting \(\bar{g} R\) one has

\[
-\bar{g} R - \frac{1}{n} \mathbb{E} R = I + \mathbb{E} [gR] + \frac{1}{n} \mathbb{E} R(\kappa D + \gamma R^\top).
\]

Finally, multiplying by \(-\frac{1}{n}\) and taking the trace,

\[
\bar{g}^2 + z \bar{g} + 1 = -\mathbb{E} g - \frac{1}{n} \mathbb{E} \text{Tr}(R(\kappa D + \gamma R^\top)) = -\mathbb{E} g^2 - \frac{1}{n^2} \mathbb{E} \text{Tr}(R(\kappa D + \gamma R^\top)). \quad (5.2.4)
\]

As a function of the entries of \(X\), \(g\) has Lipschitz constant \(O(n^{-1} \mathfrak{m}(z)^{-2})\). This fact follows from Corollary 3.8 applied to \(f(x) = 1/(x - z)\). Since the entries of \(X\) satisfy a Poincaré inequality, a standard concentration bound implies

\[
\mathbb{E} |g|^2 = O(n^{-2} \mathfrak{m}(z)^{-4}). \quad (5.2.5)
\]
Also, since $|\text{Tr}(AB)| \leq n\|A\|\|B\|$, we find

$$\left|\text{Tr}(\kappa D + \gamma R^\top)\right| = O(n\Im(z)^{-2}).$$

We deduce

$$\mathbb{E}g_2^2 = O(n^{-2}\Im(z)^{-4}) \quad \text{and} \quad \frac{1}{n^2}\mathbb{E}\text{Tr}[R(\kappa D + \gamma R^\top)] = O(n^{-1}\Im(z)^{-2}).$$

We thus have proved that

$$y^2 + zy + 1 = O_z(n^{-1}). \quad (5.2.6)$$

**Lemma 5.2.** Let $\delta \in \mathbb{C}$ and $z \in \mathbb{C}_+$. If $x \in \mathbb{C}_+$ satisfies $x^2 + zx + 1 = \delta$, then,

$$|x - g_{sc}(z)| \leq \frac{|\delta|}{\Im(z)}.$$

**Proof.** Recall that $g_{sc}^2(z) + zg_{sc}(z) + 1 = 0$. It follows that

$$\left(g_{sc}(z) + \frac{z}{2}\right)^2 = -1 + \frac{z^2}{4} \quad \text{and} \quad \left(x + \frac{z}{2}\right)^2 = -1 + \frac{z^2}{4} + \delta.$$

Hence

$$\delta = \left(\left(x + \frac{z}{2}\right)^2 - \left(g_{sc}(z) + \frac{z}{2}\right)^2\right) = (x - g_{sc}(z))(x + g_{sc}(z) + z).$$

It yields,

$$|x - g_{sc}(z)| = \frac{|\delta|}{|x + g_{sc}(z) + z|}.$$

Since $x, g_{sc}(z) \in \mathbb{C}_+, |x + g_{sc}(z) + z| \geq \Im(x + g_{sc}(z) + z) \geq \Im(z)$. \qed

From (5.2.6) and lemma 5.2, we deduce a new proof of Wigner’s semi-circular law for Gaussian Wigner matrices.

### 5.2.1 Convergence of edge eigenvalues for Gaussian matrices

We pursue the analysis of Gaussian Wigner matrices to the study of extremal eigenvalues of $X/\sqrt{n}$.

Our aim is to prove

**Theorem 5.3.** Let $X = X(n)$ be a Gaussian Wigner matrix satisfying assumptions (A). We have a.s.

$$\lim_{n \to \infty} \lambda_1 \left(\frac{X}{\sqrt{n}}\right) = -\lim_{n \to \infty} \lambda_n \left(\frac{X}{\sqrt{n}}\right) = 2.$$

A proof of this result can be proved by using the moment method, see Füredi and Komlós [FK81]. Here, we will instead use the resolvent differentiation formulas. In the GUE case, this approach was initiated by Haagerup and Thorbjørnsen [HT05] and developed notably in Schultz [Sch05], Capitaine and Donati-Martin [CDM07].

For simplicity, we set $\lambda_k = \lambda_k(X/\sqrt{n})$. Note that $X$ and $-X$ have the same law. Hence, by symmetry, we may restrict to $\lambda_1$. Note also that Wigner semi-circular law implies that a.s.

$$\lim_{n \to \infty} \inf \lambda_1 \geq 2.$$

We start with a weak bound on the norm of random matrices.
Lemma 5.4. Let $X = (X_{ij})_{1 \leq i,j \leq n} \in \mathcal{H}_n(\mathbb{C})$ be a Wigner matrix with $\mathbb{E}X_{ij} = 0$, $\mathbb{E}|X_{ij}|^2 = 1$ and $\mathbb{E}|X_{ij}|^4 \leq \kappa$. Then, there exists $c > 0$ depending only on $\kappa$ such that for any $n \geq 1$,

$$
\mathbb{E}\|X\| \leq c\sqrt{n \log n}.
$$

Proof. For $i < j$, define $Y_{ij} = X_{ij}(e_i^*e_j + e_j^*e_i)$ and $Y_{ii} = X_{ii}e_i^*e_i$. The matrices $(Y_{ij})_{i \leq j}$ are independent, centered and $X = \sum_{i \leq j} Y_{ij}$. By symmetrization, if $(\varepsilon_{ij})_{i \leq j}$ is i.i.d. symmetric Bernoulli variables independent of $X$,

$$
\mathbb{E}\|X\| \leq 2\mathbb{E}\|\sum_{i \leq j} \varepsilon_{ij}Y_{ij}\|.
$$

From Rudelson’s inequality (see Oliveira [Oli10] for a short proof), for some constant $C > 0$,

$$
\mathbb{E}\left\| \sum_{i \leq j} \varepsilon_{ij}Y_{ij} \right\| \leq C\sqrt{\log n}\left(\mathbb{E}\max_i S_i\right)^{1/2}.
$$

Observe that $\sum_{i \leq j} Y_{ij}^2$ is a diagonal matrix whose entry $(i, i)$ is equal to $S_i = \sum_{j=1}^n |X_{ij}|^2$. From Jensen’s inequality, we get

$$
\mathbb{E}\left\| \sum_{i \leq j} \varepsilon_{ij}Y_{ij} \right\| \leq C\sqrt{\log n}\left(\mathbb{E}\max_i S_i\right)^{1/2}.
$$

By assumption, $\mathbb{E}(\sum_j (|X_{ij}|^2 - 1))^2 \leq \kappa n$. Hence, from Markov inequality and the union bound, $\mathbb{P}(\max_i S_i \geq n + t) \leq \kappa n^2 t^{-2}$. So finally,

$$
\mathbb{E}\max_i S_i \leq n + \int_0^\infty 1 \land (\kappa n^2 t^{-2}) dt = (1 + 2\sqrt{\kappa})n.
$$

It concludes the proof.

By Lemma 3.18 and Borel-Cantelli Lemma, a.s.

$$
\lim_{n \to \infty} |\lambda_1 - \mathbb{E}\lambda_1| = 0.
$$

Hence, in view of Lemma (5.4), it is thus sufficient to prove that, for any $\varepsilon > 0$, a.s. for all $n \gg 1$,

$$
1(\lambda_1 \in [2(1 + \varepsilon), 2c\sqrt{\log n}]) = 0.
$$

To this end, we fix $\varepsilon > 0$ and set

$$
K = 2c\sqrt{\log n} \quad \text{and} \quad \Delta = [2 + 2\varepsilon, K].
$$

Consider a smooth function $\varphi : \mathbb{R} \mapsto [0, 1]$ with support $[2 + \varepsilon, 2K]$ such that $\varphi(x) = 1$ on $[2 + 2\varepsilon, K]$. By Lemma 3.18 and Borel-Cantelli Lemma, a.s.

$$
\lim_{n \to \infty} |\varphi(\lambda_1) - \mathbb{E}\varphi(\lambda_1)| = 0.
$$

Assume that we manage to prove that

$$
\mathbb{E}\frac{1}{n} \sum_{k=1}^n \varphi(\lambda_k) = \mathbb{E}\int \frac{\varphi d\mu_{X/\sqrt{n}}}{\sqrt{n}} \leq \frac{1}{2n}.
$$

Then, using $1(\lambda \in \Delta) \leq \varphi(\lambda)$, we would deduce, that

$$
\mathbb{P}(\lambda_1 \in \Delta) \leq \mathbb{E}\varphi(\lambda_1) \leq n\mathbb{E}\int \varphi d\mu_{X/\sqrt{n}} \leq \frac{1}{2}.
$$
And it would yields to a.s. for \( n \gg 1 \),
\[
1(\lambda_1 \in \Delta) \leq 1/3.
\]

Hence, the indicator function is equal to 0 and \( \lambda_1 \in \Delta \). It follows that if (6.5.2) holds for any \( \varepsilon > 0 \), our Theorem 5.3 is proved.

The first step of proof is to relate \( \int \varphi d\mu \) to an integral over \( g_{\mu}(z) \). For \( C^1 \) functions \( f : \mathbb{C} \to \mathbb{C} \), we set \( \hat{\partial}f(z) = \partial_x f(z) + i\partial_y f(z) \), where \( z = x + iy \). In the next lemma \( \chi : \mathbb{R} \to \mathbb{R} \) is a compactly supported smooth function such that \( \chi(y) = 1 \) in a neighbourhood of 0.

**Lemma 5.5.** Let \( k \geq 1 \) be an integer and \( \varphi : \mathbb{R} \to \mathbb{R} \) a compactly supported \( C^{k+1} \)-function, then for any \( \mu \in \mathcal{P}(\mathbb{R}) \),
\[
\int \varphi d\mu = \frac{1}{\pi} \Re \int_{\mathbb{C}_+} \tilde{\partial}\Phi(z) g_{\mu}(z) dz,
\]
where \( \Phi(x + iy) = \sum_{\ell=0}^{k} \frac{(iy)^{\ell}}{\ell!} \varphi^{(\ell)}(x) \chi(y) \).

**Proof.** Observe that in a neighbourhood of \( \mathbb{R} \),
\[
\tilde{\partial}\Phi(x + iy) = \varphi^{(k+1)}(x)(iy)^{k}/k!.
\]
It follows that for any \( \lambda \in \mathbb{R} \), \( \tilde{\partial}\Phi(z)/(z - \lambda) \) is integrable. Now, from Fubini’s Theorem, it suffices to check this for \( \mu = \delta_0 \):
\[
\varphi(0) = \Phi(0) = -\frac{1}{\pi} \Re \int_{\mathbb{C}_+} \tilde{\partial}\Phi(z) \frac{dz}{z},
\]
If \( z = x + iy \), we have
\[
\Re \left( \frac{\tilde{\partial}\Phi(z)}{z} \right) = \frac{x \partial_x \Phi(z)}{x^2 + y^2} + \frac{y \partial_y \Phi(z)}{x^2 + y^2} = \frac{1}{r} \partial_r \Phi(z),
\]
where \( r = |z| \) and \( \partial_r \) is the radial derivative in polar coordinates. We thus find
\[
\Re \int_{\mathbb{C}_+} \tilde{\partial}\Phi(z) \frac{dz}{z} = \int_{0}^{\pi} \int_{0}^{\infty} \partial_r \Phi(re^{i\theta}) rdrd\theta = -\pi \Phi(0).
\]

**GUE case:** we can now prove Theorem 5.3 for GUE matrices. Indeed, in this case, \( \gamma = 0 \) and \( \kappa = 0 \). In particular, from Equations (5.2.4) and (5.2.5), we find
\[
\bar{g}^2 + z\bar{g} + 1 = O(n^{-2}\Im(z)^{-4}).
\]
From Lemma 5.2, we find
\[
\bar{g} - g_{sc} = \delta(z) = O(n^{-2}\Im(z)^{-5}).
\]
We may apply Lemma 5.3 with \( k = 6 \) to our smooth function \( \varphi \) with support \([2 + \varepsilon, 2K]\). We find
\[
\mathbb{E} \int \varphi d\mu_{X/\sqrt{n}} = \frac{1}{\pi} \Re \mathbb{E} \int_{2+\varepsilon}^{\infty} \int_{0}^{\infty} \tilde{\partial}\Phi(x + iy) \bar{g}(x + iy) dy dx
\]
\[
= \frac{1}{\pi} \Re \mathbb{E} \int_{2+\varepsilon}^{\infty} \int_{0}^{\infty} \tilde{\partial}\Phi(x + iy) (g_{sc}(x + iy) + \delta(x, y)) dy dx
\]
\[
= \int \varphi d\mu_{sc} + \frac{1}{\pi} \Re \mathbb{E} \int_{2+\varepsilon}^{2K} \int_{0}^{\infty} \tilde{\partial}\Phi(x + iy) \delta(x, y) dy dx
\]
Now, the support of $\mu_{sc}$ is $[-2, 2]$. Hence $\int \varphi d\mu_{sc} = 0$. Also, from (5.2.10),
\[
\partial \Phi(x + iy)\delta(x, y) = O(n^{-2}).
\]
and is compactly supported from (5.2.7) and the definition of $\varphi$. We thus have proved that
\[
\mathbb{E} \int \varphi d\mu_{X/\sqrt{n}} = O(n^{-2} \sqrt{\log n}).
\]
It concludes the proof of (6.5.2) in the GUE case.

**General Gaussian case:** the above argument cannot work directly if $\gamma \neq 0$ or $\kappa \neq 0$. Indeed, Equation (5.2.4) gives only
\[
g^2 + zg + 1 = O(n^{-1} \Im(z)^{-2} + n^{-2} \Im(z)^{-4}).
\]
and from Lemma 5.2,
\[
g = g_{sc} + O(n^{-1}(\Im(z)^{-5} \wedge 1)). \tag{5.2.11}
\]
We thus have to study more precisely
\[
\frac{1}{n^2} \text{Tr}(R(\kappa D + \gamma R^\top)).
\]
We first need a lemma

**Lemma 5.6.** For any $\varepsilon > 0$, we have
\[
\inf \{ |g_{sc}(z)| : z \in \mathbb{C}, d(z, [-2, 2]) \geq \varepsilon \} < 1,
\]
where $d(z, A) = \inf \{|w - z| : w \in A\}$.

**Proof.** Let $r(z) = |g_{sc}(z)|$ and $t = \varepsilon/\sqrt{2}$. If $d(z, [-2, 2]) \geq \varepsilon$ then either $|\Im(z)| \geq t$ or $|\Re(z)| \geq 2 + t$. We first assume that $\Im(z) \geq t$. Note that if $z = E + i\eta$ and $\xi$ has distribution $\mu_{sc}$, then $r(z) = \mathbb{E}[(\xi - E - i\eta)^{-1}]$. By symmetry and monotonity, $r(z) \geq r(i\eta) \geq r(it)$. We find
\[
r(it) = \frac{1}{2\pi} \int_{-2}^{2} \frac{t\sqrt{4 - x^2}}{t^2 + x^2} dx = \frac{1}{2\pi} \int_{-2/t}^{2/t} \frac{\sqrt{4 - (tx)^2}}{1 + x^2} dx < \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2}{1 + x^2} dx = 1.
\]
It remains to deal with $z = E + i\eta$ and $|E| \geq 2 + t$. We have $r(z) \geq r(E) = r(|E|) \geq r(2 + t)$ and
\[
r(2 + t) = \frac{1}{2\pi} \int_{-2}^{2} \frac{\sqrt{4 - x^2}}{2 + t - x} dx < \frac{1}{2\pi} \int_{-2}^{2} \frac{\sqrt{4 - x^2}}{2 - x} dx = 1.
\]

**Lemma 5.7.** Let $f(z) = \gamma g^2_{sc}(z)/(\gamma g^2_{sc}(z) + 1) + \kappa g^2_{sc}(z)$, we have
\[
\mathbb{E} \frac{1}{n} \text{Tr}(R(\kappa D + \gamma R^\top)) = f(z) + O(n^{-1}(1 - |g_{sc}(z)|^2)^{-1}(1 \wedge \Im(z)^{-5})).
\]

**Proof.** We may again use the Gaussian integration by part formula. Using (5.2.1)-(5.2.2)-(5.2.3), we get, for $0 \leq a \neq b \leq n$, and all $j, k, \ell, m$:
\[
\mathbb{E}R_{jk}R_{\ell m}X_{ab} = -\frac{1}{\sqrt{n}} (\mathbb{E}(\gamma R_{ja}R_{bk} + R_{jb}R_{ak})R_{\ell m} + \mathbb{E}R_{jk}(\gamma R_{ta}R_{bm} + R_{tb}R_{am})), \tag{5.2.12}
\]
and
\[
\mathbb{E}R_{jk}R_{\ell m}X_{aa} = -\frac{\sigma^2}{\sqrt{n}} (\mathbb{E}R_{ja}R_{ak}R_{\ell m} + \mathbb{E}R_{jk}R_{ta}R_{am}), \tag{5.2.13}
\]
5.2. GAUSSIAN DIFFERENTIATION FORMULAS

We use again the identity $-zR = I - \frac{1}{\sqrt{n}}RX$. Taking conjugate and composing by $R$ yields to $-zRR^T = R - \frac{1}{\sqrt{n}}RX^TR^T$, we find

$$-z(RR^T)_{kk} = R_{kk} - \frac{1}{\sqrt{n}} \sum_{a,b} R_{ka}R_{kb}X_{ab}.$$ 

We now take expectation and use (5.2.12)-(5.2.13),

$$\mathbb{E}(RR^T)_{kk} = \mathbb{E}R_{kk} + \frac{2}{n} \sum_{a \neq b} \mathbb{E}R_{ka}^2R_{bb} + \frac{1}{n} \sum_{a \neq b} \mathbb{E}R_{kb}R_{aa}R_{ka} + \frac{2}{n} \sum_{a \neq b} \mathbb{E}R_{ka}R_{ka}R_{ba} + \frac{1}{n} \sum_{a \neq b} \mathbb{E}R_{ka}^2R_{aa} + \frac{2\sigma^2}{n} \sum_{a} \mathbb{E}R_{ka}^2R_{aa}.$$ 

We set

$$m = \frac{1}{n} \text{Tr}(RR^T), \quad m = \mathbb{E}m \quad \text{and} \quad \overline{m} = m - \mathbb{E}m.$$ 

Recall that $D_{kk} = R_{kk}$. Taking Tr and dividing by $n$, in the above expression we obtain

$$-z\overline{m} = \overline{g} + (\gamma + 1)\mathbb{E}gm + \frac{1}{n^2} \text{Tr}(R(R^T)^2 + \gamma R^2R^T + 2\kappa RR^T D).$$ 

We deduce that

$$-(z + (\gamma + 1)\overline{g})\overline{m} = \overline{g} + \mathbb{E}gm + \frac{1}{n^2} \text{Tr}(RR^T((1 + \gamma)R^T + 2\kappa D)).$$ 

Using (5.2.5), $|m| \leq \Im(z)^{-2}$ and $|\text{Tr}(A)| \leq n||A||$, we find

$$-(z + (\gamma + 1)\overline{g})\overline{m} = \overline{g} + O(n^{-1}\Im(z)^{-4} \land 1)).$$ 

We deduce from (5.2.11)

$$-(z + (\gamma + 1)g_{sc})\overline{m} = g_{sc} + O(n^{-1}\Im(z)^{-5} \land 1)).$$ 

We multiply by $g_{sc} = O(1)$ and use that $g_{sc}^2 + zg_{sc} + 1 = 0$. From Lemma 5.6 and $|\gamma| \leq 1, |\gamma g_{sc}^2 + 1| \geq 1 - |g_{sc}|^2 > 0$. We find

$$\overline{m} = \frac{g_{sc}^2}{g_{sc}^2 + 1} + O(n^{-1}(1 - |g_{sc}(z)|^2)^{-1}(\Im(z)^{-5} \land 1)).$$ 

we have $(z + (\gamma + 1)g_{sc})$.

We set similarly

$$m' = \frac{1}{n} \sum_{k=1}^n (R_{kk})^2, \quad m' = \mathbb{E}m' \quad \text{and} \quad \overline{m'} = m' - \mathbb{E}m',$$

so that $\frac{1}{\sqrt{\gamma}} \text{Tr}(RD) = m'$. From $-zR = I - \frac{1}{\sqrt{n}}RX$, multiplying by $R_{kk}$, we obtain

$$-z(R_{kk})^2 = R_{kk} - \frac{1}{\sqrt{n}} \sum_a R_{ka}R_{kk}X_{ak}.$$ 

We find analogously

$$-(z + \overline{g})m' = \overline{g} + O(n^{-1}\Im(z)^{-3}).$$ 

and, from (5.2.11), $g_{sc}^2 + zg_{sc} + 1 = 0$ and $|g_{sc}| = O(1)$.

$$m' = g_{sc}^2 + O(n^{-1}\Im(z)^{-5} \land 1)).$$

This concludes the proof.
CHAPTER 5. RESOLVENT DIFFERENTIATION

We may now conclude the proof of Theorem 6.5.2. Let $K = \{ z \in \mathbb{C} : \Re(z) \geq 2 + \varepsilon \}$ and $K_+ = K \cap \mathbb{C}_+$. We also set

$$L = \frac{1}{zn^2} \text{Tr}[R(\kappa D + \gamma R^*)].$$

On $K_+$, we have $L = O(n^{-1}\Im(z)^{-2})$. From Equations (5.2.4) and (5.2.5), we find

$$(g + L)^2 + z(g + L) + 1 = O(n^{-2}\Im(z)^{-4}).$$

For $n$ large enough, $g + L \in \mathbb{C}_+$. Hence, from Lemma 5.2, we find

$$g + L - g_{sc} = \delta(z) = O(n^{-2}\Im(z)^{-5}).$$

So finally, from (5.2.11) and Lemma 5.7, for all $z \in K_+$,

$$g = g_{sc} - f(z) + O(n^{-2}(\Im(z)^{-5} \wedge 1)), \quad (5.2.16)$$

where the $O(\cdot)$ depends on $\varepsilon$.

As above, for our smooth function $\varphi$ with support $[2 + \varepsilon, 2K]$ we may apply Lemma 5.5 with $k = 5$. We find

$$\mathbb{E} \int \varphi d\mu_{X/\sqrt{n}} = \frac{1}{\pi} \text{Re} \int_{K_+} \partial \Phi(x + iy) \overline{\varphi}(x + iy) dx dy$$

$$= \frac{1}{\pi} \text{Re} \int_{K_+} \partial \Phi(x + iy)(g_{sc}(x + iy) - f(x + iy))/n + \delta(x, y)) dx dy$$

$$= \frac{1}{\pi} \text{Re} \int_{K_+} \partial \Phi(x + iy)(g_{sc}(x + iy) - f(x + iy))/n dx dy$$

$$= \frac{1}{\pi} \text{Re} \int_{K_+} \partial \Phi(x + iy)\delta(x, y) dx dy$$

Now, notice that $g_{sc}(z) - f(z)/n$ is analytic on an open neighbourhood of $K_+$. In particular, $\partial(g_{sc} - f/n) = 0$ on this neighbourhood. Hence, by integration by part, the first integral of the above expression is 0. Also, from (5.2.10)

$$\partial \Phi(x + iy)\delta(x, y) = O(n^{-2}).$$

and is compactly supported. We thus have proved that

$$\mathbb{E} \int \varphi d\mu_{X/\sqrt{n}} = O(n^{-2}\sqrt{\log n}).$$

It concludes the proof of Theorem 5.3.

5.2.2 Convergence of edge eigenvalues under $4 + \varepsilon$ moment condition

In this paragraph, we consider a new set of assumptions on the Wigner matrix $X = (X_{ij})_{1 \leq i, j \leq n}$. We assume that the probability distributions $P$ and $Q$ of $X_{12}$ and $X_{11}$ satisfy

(B1) $\mathbb{E}X_{12} = 0$, $\mathbb{E}|X_{12}|^2 = 1$ and $\mathbb{E}|X_{12}|^p < \infty$ for some $p > 4$.

(B2) $\mathbb{E}X_{11} = 0$ and $\mathbb{E}|X_{11}|^2 < \infty$.

Theorem 5.8. Let $X = X(n)$ be a Wigner matrix satisfying assumptions (B). We have a.s.

$$\lim_{n \to \infty} \lambda_1 \left( \frac{X}{\sqrt{n}} \right) = -\lim_{n \to \infty} \lambda_n \left( \frac{X}{\sqrt{n}} \right) = 2.$$
5.2. GAUSSIAN DIFFERENTIATION FORMULAS

The proof of Theorem 5.8 will follow the same strategy than the proof for Gaussian Wigner matrices. There will be however two new difficulties: matrices satisfying assumptions (B) do not have (i) as strong concentration inequalities and (ii) the Gaussian integration by part is no longer exactly available. We will solve issue (i) by a truncation step and the use of Talagrand’s concentration inequality. Issue (ii) will be addressed with the following lemma due to Khorunzhy, Khoruzhenko and Pastur [KKP96].

Lemma 5.9. Let \( \xi \) be a real-valued random variable such that \( \mathbb{E}|\xi|^{k+2} < \infty \). Let \( f : \mathbb{R} \to \mathbb{C} \) be a \( C^{k+1} \) function such that the \((k+1)\)-th derivative is uniformly bounded. Then,

\[
\mathbb{E}\xi f(\xi) = \sum_{\ell=0}^{k} \frac{k_{\ell+1}}{\ell!} \mathbb{E}f^{(\ell)}(\xi) + O(\|f^{(k+1)}\|_{\infty}\mathbb{E}|\xi|^{k+2}),
\]

where \( k_{\ell} \) is the \( \ell \)-th cumulant of \( \xi \) and \( O(\cdot) \) depends only on \( k \).

Proof. \( \cdots \)

Truncation: We start the proof of Theorem 5.8 by the truncation step. Consider the new set of assumptions:

(C1) \( \mathbb{E}X_{12} = 0, \mathbb{E}|X_{12}|^2 = 1, \mathbb{E}|X_{12}|^4 < C \) and \( \text{supp}(P) \subset B(0, n^{1/q}) \).

(C2) \( Q(0) = 1 \).

Recall that \( P \) is the common distribution of \( (X_{ij})_{1 \leq i < j \leq n} \) and \( Q \) is the common distribution of \( (X_{ii})_{1 \leq i \leq n} \). Note here that \( P \) depends on \( n \) and we shall write the variables as \( X_{ij}(n) \) and assume that they are on a common probability space.

The next lemma is due to Bai and Yin [BY88].

Lemma 5.10 (Truncation). If the conclusion of Theorem 5.8 holds under assumptions (C) then it also holds under assumptions (B).

Proof. We first get rid of the diagonal terms. Recall that, for a non-negative random variable \( Y \) and \( a > 0 \),

\[
\mathbb{E}Y^a = a \int_0^\infty t^{a-1}\mathbb{P}(Y > t)dt \geq k \sum_{\ell=1}^{\infty} a_{\ell+1}^a(u_{\ell+1} - u_{\ell})\mathbb{P}(Y > u_{\ell}), \tag{5.2.17}
\]

where \( (u_{\ell}) \) is any non-negative increasing sequence. Hence, assumption (B2) implies that for any \( \varepsilon > 0 \),

\[
\sum_{\ell=1}^{\infty} \varepsilon^2 \mathbb{P}(|X_{11}| > \varepsilon 2^{\ell/2}) < \infty.
\]

In particular, the above holds for some sequence \( \varepsilon = \varepsilon(n) \) going to 0. Let \( E_n \) be the event that \( \max_{1 \leq i \leq n} |X_{ii}| \leq \varepsilon \sqrt{2n} \). We claim that a.s. for all \( n \gg 1 \), \( E_n \) holds. Indeed,

\[
\sum_{k \geq 1} \mathbb{P}\left( \bigcup_{2^{k-1} \leq n < 2^k} E_n \right) = \sum_{k \geq 1} \mathbb{P}\left( \bigcup_{2^{k-1} \leq n < 2^k} \bigcup_{1 \leq i < n} \{ |X_{ii}| > \varepsilon \sqrt{2n} \} \right) \\
\leq \sum_{k \geq 1} \mathbb{P}\left( \bigcup_{1 \leq i < n} |X_{ii}| > \varepsilon 2^{k/2} \right) \\
\leq \sum_{k \geq 1} 2^k \mathbb{P}(|X_{11}| > \varepsilon 2^{k/2}).
\]
It follows, from Borel-Cantelli’s lemma, that a.s. for \( n \gg 1 \), \( E_n \) holds.

If \( \hat{X} = X - \text{diag}(X) \) is the matrix obtained by removing the matrix with the diagonal entries, we have, if \( E_n \) holds,

\[
\|X\| - \|\hat{X}\| \leq \|\text{diag}(X)\| \leq \varepsilon\sqrt{2n} = o(\sqrt{n}).
\]

We have thus checked that we may assume without loss of generality that (C2) holds.

Similarly, if \( F_n \) denotes the event that \( \max_{1 \leq i, j \leq n} |X_{ij}| \leq \varepsilon(2n)^{2/p} \), we have

\[
\sum_{n \geq 1} \mathbb{P}(F_n^c) = \sum_{k \geq 1} \mathbb{P}\left( \bigcup_{2^k-1 \leq n < 2^k} \bigcup_{1 \leq i < j \leq n} \left\{ |X_{ij}| > \varepsilon(2n)^{2/p} \right\} \right)
\leq \sum_{k \geq 1} \mathbb{P}\left( \bigcup_{1 \leq i < j \leq 2^k} \left\{ |X_{ij}| > \varepsilon 2^{2k/p} \right\} \right)
\leq \sum_{k \geq 1} 2^{2k} \mathbb{P}(|X_{12}| > \varepsilon 2^{2k/p})
\]

which is finite from (5.2.17) applied to \( a = p \) and \( u = 2^a/p \). We have checked so far that it is sufficient to prove Theorem 5.8 for the matrix \( Y = (Y_{ij})_{1 \leq i, j \leq n} \) with \( Y_{ij} = X_{ij}1(|X_{ij}| \leq \varepsilon n^{1/q}) \) (and \( X_{ii} = 0 \)) with \( q = p/2 \). The variable \( Y_{12} \) is however a priori no longer centered with unit variance. If \( \rho(n) = \mathbb{E}X_{12}1_{|X_{12}| \leq \varepsilon n^{1/q}} = -\mathbb{E}X_{12}1_{|X_{12}| > \varepsilon n^{1/q}} \), we find, from Hölder inequality,

\[
|\rho(n)| \leq (\mathbb{E}|X_{11}|^p)^{1/p} \left( \mathbb{P}(|X_{11}| \geq \varepsilon n^{1/q}) \right)^{1-1/p} = O(\varepsilon^{1-p}n^{-3/2}).
\]

If \( \hat{Y} \) has entries \( \hat{Y}_{ij} = 1(i \neq j)(Y_{ij} - \rho(n)) \),

\[
\left\| Y \right\| - \left\| \hat{Y} \right\| \leq \left\| Y - \hat{Y} \right\| = \left\| (J - I)\rho(n) \right\| \leq n\rho(n) = o(1).
\]

Similarly, if \( \sigma^2(n) = \text{Var}(\hat{Y}_{12}) = \mathbb{E}|X_{12}1(|X_{12}| \leq \varepsilon n^{1/q}) - \rho(n)|^2 \), by dominated convergence, we find \( \lim_{n \to \infty} \sigma^2(n) = 1 \). If follows that

\[
\left\| \frac{\hat{Y}}{\sigma \sqrt{n}} \right\| - \left\| \frac{\hat{Y}}{\sqrt{n}} \right\| = \left\| \frac{\hat{Y}}{\sigma \sqrt{n}} \right\| - \left\| \frac{\hat{Y}}{\sqrt{n}} \right\| = \left\| \frac{\hat{Y}}{\sigma \sqrt{n}} \right\| - \left\| \frac{\hat{Y}}{\sqrt{n}} \right\| = o(\left\| \frac{\hat{Y}}{\sqrt{n}} \right\|).
\]

It concludes the proof since the matrix \( \hat{Y}/\sigma \) has now all required properties.

**Convergence in probability:** Assume that assumptions (C) holds. The aim of this paragraph is to prove that for any \( \varepsilon > 0 \), the following holds

\[
\mathbb{P}\left( \lambda_1\left( \frac{X}{\sqrt{n}} \right) \in \Delta \right) = o(1),
\]

(5.2.18)

(in words : the largest eigenvalues of \( X/\sqrt{n} \) converges in probability to 2). The central part of the proof will be in the next lemma.

**Lemma 5.11.** Under assumptions (C), if \( \delta = (1/2 - 1/q) \land 1/2 > 0 \),

\[
\overline{g}(z) = g_{sc}(z) - \frac{f(z)}{n} + O\left( (1 - |g_{sc}(z)|^2)^{-1}n^{-1-\delta}(1 \land \Im(z)^{-6}) \right).
\]

where \( f \) is given by Lemma 5.7 with \( \gamma = \mathbb{E} X_{12}^2 \) and \( \sigma^2 = 0 \).
5.2. GAUSSIAN DIFFERENTIATION FORMULAS

Notice that $\gamma$ and $f$ depend on $n$. Before proving Lemma 5.11, let us check that (5.2.18) holds. We observe that $|\gamma| \leq E|X_{12}|^2 = 1$ implies that $f$ is analytic on $[2 + \varepsilon, \infty) \times [0, \infty)$. Then, the argument below (5.2.16) with $k = 6$ and $\delta(x,y) = O((1 - |g_{sc}(x + iy)|^2)^{-1}n^{-1-\delta}y^{-\varepsilon})$ gives

$$E \int \varphi \mu_X/\sqrt{n} = O(n^{-1-\delta} \sqrt{\log n}).$$

(5.2.19)

Then (5.2.18) follows directly from (5.2.9).

We now turn to the proof of Lemma 5.11.

Proof of Lemma 5.11. The computation leading to (5.2.4) and Lemma 5.9 with $k = 3$ give

$$g^2 + zg + 1 = -Eg^2 - \frac{1}{n^2}ETr[R(\kappa D + \gamma R^T)] + \frac{\kappa_3}{n^{5/2}}\Omega_2 + \frac{\kappa_4}{n^3}\Omega_3 + \frac{E|X_{12}|^5}{n^{7/2}}\Omega_4,$$

(5.2.20)

where $\Omega_2$ and $\Omega_3$ is a double sum of second and third derivatives of resolvent entries and $\Omega_4$ is bounded up to a multiplicative constant by a double sum of sup of fourth derivatives.

Let us start with the last term. If $G = (X - z)^{-1}$, notice that $\partial^4_{ab}G_{k\ell}$ is a finite sum of terms in

$$G_{k_{\varepsilon_1}}G_{\varepsilon_1\varepsilon_2}G_{\varepsilon_2\varepsilon_3}G_{\varepsilon_3\varepsilon_4}G_{\varepsilon_4\ell},$$

where $\varepsilon_i \in \{a,b\}, \bar{a} = b$ and $\bar{b} = a$. It follows that $\Omega_4$ is bounded up to a multiplicative constant by

$$\sum_{\varepsilon_i \in \{a,b\}} \sum_{1 \leq k,a \leq n} |R_{k\varepsilon_1}R_{\varepsilon_1\varepsilon_2}R_{\varepsilon_2\varepsilon_3}R_{\varepsilon_3\varepsilon_4}R_{\varepsilon_4\ell}|.$$

Since $E|X_{12}|^5 \leq n^{1/4}E|X_{12}|^4 = O(n^{1/2-\delta})$ and $|R_{jk}| \leq \Im m(z)^{-1}$, we deduce that

$$\frac{E|X_{12}|^5}{n^{7/2}}\Omega_4 = O\left(\frac{n^{1/2-\delta}n^2\Im m(z)^{-5}}{n^{7/2}}\right) = O\left(n^{-1-\delta}\Im m(z)^{-5}\right).$$

(5.2.21)

Let us now deal with $\Omega_2$. For some constants $c_\varepsilon$ which are not necessary to compute, we have

$$\Omega_2 = \sum_{\varepsilon = (a,a)} c_\varepsilon \sum_{1 \leq k,a \leq n} R_{k\varepsilon_1}R_{\varepsilon_1\varepsilon_2}R_{\varepsilon_2a}.$$

If $\varepsilon = (a,a)$ we have

$$\left|\sum_{k,a} (R_{ka})^3\right| \leq \Im m(z)^{-1} \sum_{k,a} |R_{ka}|^2 = \Im m(z)^{-1}Tr(RR^*) \leq n\Im m(z)^{-3}.$$

To deal with the case $\varepsilon = (a,k)$, we introduce the vector $x = (x_k)$ with $x_k = R_{kk}$ and the matrix $M_{kk} = R_{kk}R_{kk}$. We find

$$\left|\sum_{k,a} R_{ka}R_{kk}R_{aa}\right| = \sum_k (Mx)_k \leq \sqrt{n}||Mx|| \leq \sqrt{n}||M||||x||.$$

Now, notice that $M = DR$. Since $||R|| \leq \Im m(z)^{-1}$ and $|R_{kk}| \leq \Im m(z)^{-1}$, it implies that $||M|| \leq \Im m(z)^{-2}$ and $||x|| \leq \sqrt{n}\Im m(z)^{-1}$ so finally

$$\left|\sum_{k,a} R_{ka}R_{kk}R_{aa}\right| \leq n\Im m(z)^{-3}.$$
To deal with the cases $\varepsilon = (k, k)$ and $\varepsilon = (k, a)$, we observe that

$$\sum_{k,a} R_{kk}R_{ak}R_{aa} = \sum_{k,a} R_{kk}R_{aa}R_{ka} = \sum_{k,a} R_{ka}R_{kk}R_{aa}. $$

So finally,

$$\frac{K_3}{n^{3/2}} \Omega_2 = O\left(n^{-3/2}|\mathfrak{m}(z)|^{-3}\right). \quad (5.2.22)$$

We may repeat the same argument for $\Omega_3$. For some new constants $c_\varepsilon$,

$$\Omega_3 = \sum_{\varepsilon=(\varepsilon_1,\varepsilon_2,\varepsilon_3)\in\{a,k\}^3} c_\varepsilon \sum_{1 \leq k,a \leq n} R_{k\varepsilon_1}R_{\varepsilon_1\varepsilon_2}R_{\varepsilon_2\varepsilon_3}R_{\varepsilon_3a}. $$

The expression

$$S(\varepsilon) = \sum_{1 \leq k,a \leq n} R_{k\varepsilon_1}R_{\varepsilon_1\varepsilon_2}R_{\varepsilon_2\varepsilon_3}R_{\varepsilon_3a}. $$

is invariant by a cyclic permutation of $(\varepsilon_1,\varepsilon_2,\varepsilon_3)$. We also find that $S(a,a,k) = S(k,k,k)$. There has thus three cases to consider. We get

$$S(a,a,a) = \sum_{1 \leq k,a \leq n} (R_{ka})^4 \leq |\mathfrak{m}(z)|^{-2}\text{Tr}(RR^*) \leq n|\mathfrak{m}(z)|^{-4}, $$

and

$$S(a,a,k) = \sum_{1 \leq k,a \leq n} R_{kk}R_{ka}^2R_{aa} = \text{Tr}(DRDR^T). $$

In particular $|S(a,a,k)| \leq n|\mathfrak{m}(z)|^{-4}$. Similarly,

$$S(k,k,a) = \sum_{1 \leq k,a \leq n} R_{kk}R_{ak}R_{aa}R_{ka} = \text{Tr}(DRDR). $$

It follows that

$$\frac{K_4}{n^3} \Omega_3 = O\left(n^{-2}|\mathfrak{m}(z)|^{-4}\right). \quad (5.2.23)$$

We now turn to the first term on the right hand side of [5.2.20]. If $z = E + i\eta$, we have

$$\frac{1}{x-z} = \frac{(x-E)}{(x-E)^2 + \eta^2} + \frac{i\eta}{(x-E)^2 + \eta^2} = \eta^{-1}h_1\left(\frac{x-E}{\eta}\right) + i\eta^{-1}h_2\left(\frac{x-E}{\eta}\right). $$

The function $h_2$ is convex and the second derivative of $h_1$ being

$$h_1''(x) = \frac{2x(x^2 - 3)}{(1+x^2)^2}, $$

$h_1$ has 3 inflection points. From Corollary 3.22 we deduce that

$$\mathbb{E}|\mathbf{g}|^2 = O\left(|\mathfrak{m}(z)|^{-4}n^{-2}(n^{1/2-\delta})^2\right) = O\left(|\mathfrak{m}(z)|^{-4}n^{-1-2\delta}\right). \quad (5.2.24)$$

It finally remains to take care of the second term in [5.2.20]. We have the rough bound $\text{Tr}(R(\kappa D + \gamma R^T)) = O(n|\mathfrak{m}(z)|^{-2})$. Using Lemma 5.2 we have proved so far that

$$\mathbf{g} = g_{sc} + O(n^{-1-\delta}(1 \wedge |\mathfrak{m}(z)|^{-5}). \quad (5.2.25)$$

The computation leading to [5.2.14] shows that

$$-(z + (\gamma + 1)\mathbf{g})\overline{\mathbf{m}} = \overline{\mathbf{g}} + \mathbb{E}g_{m} + \frac{1}{n^2}\text{Tr}(RR^T((1+\gamma)R^T + 2\kappa D))$$

$$+ \frac{K_3}{n^{3/2}} \Omega_2^{(2)} + \frac{K_4}{n^3} \Omega_3^{(2)} + \frac{\mathbb{E}|X_{12}|^5}{n^{7/2}} \Omega_4^{(2)},$$
where, \( \Omega'_k \) is a double sum of \( k \)-th derivatives of \( RR^T \). Before estimating these terms, we use (5.2.25), (5.2.24) and \(|m| \leq \Im(z)^{-2} \). We find

\[
-(z+(\gamma+1)g_{sc})\overline{m} = g_{sc} + O(n^{-1/2}(1 \wedge \Im(z)^{-5})) + \frac{\kappa_3}{n^{5/2}} \Omega_2^{(2)} + \frac{\kappa_4}{n^3} \Omega_3^{(2)} + \frac{\Im[X_{12}]^5}{n^{7/2}} \Omega_4^{(2)},
\]  

(5.2.26)

Similarly, the computation leading to (5.2.15) gives

\[
\overline{m}' = g_{sc} + O(n^{-1/2}(1 \wedge \Im(z)^{-5})) + \frac{\kappa_3}{n^{5/2}} \Omega_2^{(3)} + \frac{\kappa_4}{n^3} \Omega_3^{(3)} + \frac{\Im[X_{12}]^5}{n^{7/2}} \Omega_4^{(3)},
\]  

(5.2.27)

where, \( \Omega_2^{(3)} \) is a double sum of \( k \)-th derivatives of \( RD \). The analysis of \( \Omega_2^{(2)} \), \( \Omega_3^{(2)} \) parallels the analysis of \( \Omega_2 \). If \( G^{(i)} \) is equal to either \( RR^T \) or \( RD \) depending on \( i = 2 \) or \( i = 3 \). We have

\[
\Omega_2^{(i)} = \sum_{\epsilon=(\epsilon_1, \epsilon_2) \in \{a, b\}^2} c_{\epsilon}^{(i)} \sum_{1 \leq k, a \leq n} G_{ka}^{(i)} R_{\epsilon_1 a} R_{\epsilon_2 a},
\]

and

\[
\Omega_3^{(i)} = \sum_{\epsilon=(\epsilon_1, \epsilon_2, \epsilon_3) \in \{a, b\}^3} c_{\epsilon}^{(i)} \sum_{1 \leq k, a \leq n} G_{ka}^{(i)} R_{\epsilon_1 a} R_{\epsilon_2 a} R_{\epsilon_3 a}.
\]

Since \( \|G^{(i)}\| \leq \Im(z)^{-2} \), all terms can be bounded as we did previously for \( \Omega_2 \). It suffices to multiply all previous bounds by \( \Im(z)^{-1} \). For example, in \( \Omega_2^{(i)} \), for \( \epsilon = (a, a) \), we write,

\[
\left| \sum_{1 \leq k, a \leq n} G_{ka}^{(i)} R_{ka} \right| \leq \Im(z)^{-2} \sum_{1 \leq k, a \leq n} |R_{ka}|^2 = \Im(z)^{-2} n \Tr(RR^*) \leq n \Im(z)^{-4}.
\]

The remainder \( \Omega_3^{(2)} \) is now a triple sum of terms of the form \( \prod_{\ell=1}^{6} R_{\epsilon_\ell \epsilon_{\ell+1}} \) while \( \Omega_3^{(3)} \) remains a double sum of such terms. In any case, we find

\[
|\Omega_3^{(i)}| \leq n^3 \Im(z)^{-6}.
\]

Recall that \( \Im[X_{12}]^5 = O(n^{1/2-\delta}) \). Hence, plugging these bounds into (5.2.26)-(5.2.27) yields to

\[
\frac{1}{n} \Im[X_{12}]^5 = f(z) + O((1-|g_{sc}|^2)^{-1} n^{-\delta} \Im(z)^{-6}).
\]

The lemma is proved. \( \square \)

**Convergence almost sure:** It follows however from Theorem 3.19, Corollary 3.8 and the convexity of \( M \mapsto \lambda_1(M) \) in \( \mathcal{H}_n(\mathbb{C}) \) (which follows from Courant-Fischer min-max Theorem 3.1), that for some constants \( c, \rho > 0 \),

\[
\mathbb{P}\left( \left| \lambda_1\left( \frac{X}{\sqrt{n}} \right) - \rho \right| \geq t \right) \leq c^{-1} \exp(-cn^{1-4/\rho} t^2).
\]

In particular, from Borel-Cantelli’s Lemma, it suffices to prove that \( \rho \leq 2(1+\varepsilon) \). First, from Lemma 5.4 we find \( \rho \leq K \). We then need to check that \( \rho \notin \Delta \), the latter follows from (5.2.19).
Chapter 6

Covariance matrices

6.1 Model description

Let $Y$ be a random vector in $\mathbb{R}^n$. Consider $(X_i)_{1 \leq i \leq N}$ independent and identically distributed copies of $Y$. We define

$$S = \frac{n}{N} \sum_{i=1}^{N} X_iX_i^t$$

(6.1.1)

We are interested by the spectrum of $S$ as $n,N \to \infty$. We will assume that $(n,N)$ are such that

$$c = \frac{n}{N} \in (C^{-1}, C),$$

for some $C > 1$. (The dependence in $(n,N)$ of $c$ is implicit). Our constants will depend on $C$.

We consider the following set of statistical assumptions.

(A) $Y$ is centered and isotropic:

$$\mathbb{E}[Y] = 0 \quad \text{and} \quad \mathbb{E}[YY^t] = I_n.$$

(B) For some $\varepsilon_n \to 0$, for any $A \in \mathcal{M}_n(\mathbb{R})$,

$$\mathbb{E}\left(Y^tAY - \frac{1}{n} \text{Tr}A\right)^2 \leq \|A\|^2 \varepsilon_n^2 = o(\|A\|^2).$$

(C) $Y$ has i.i.d. coordinates and for some $p > 4$, $\mathbb{E}(\sqrt{n}Y_1)^p < C$.

From Lemma 4.9 if $Y = (y_1, \cdots, y_n)/\sqrt{n}$ has i.i.d. coordinates and the law of $y_i$ has finite fourth moment and is independent of $n$, (B) holds with $\varepsilon_n = O(1/\sqrt{n})$.

We denote by $0 \leq \lambda_n \leq \cdots \leq \lambda_1$ the eigenvalues of $S$. Assumption (A) guarantees that $M$ is properly normalized:

$$\frac{1}{n} \mathbb{E} \text{Tr}S = \mathbb{E} \frac{1}{n} \sum_{i=1}^{n} \lambda_i = 1.$$ 

In this chapter, our aim is to estimate

$$\|S - I\| = \max(1 - \lambda_n, \lambda_1 - 1),$$

as $n, N \to \infty$.

As usual, the empirical spectral distribution (ESD) of $M$ is

$$\mu_S = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}.$$
It has to be compared with the Marchenko-Pastur distribution with parameter $c$

$$
\mu_c = \frac{1}{2\pi c \epsilon} \sqrt{(b-x)(x-a)} 1_{x \in [a, b]} dx + (c - 1) \delta_0
$$

where $a = ((1 - \sqrt{c})^2, b = (1 + \sqrt{c})^2$.

We denote by $d(\mu, \nu)$ the Lévy distance between probability measures. Our first statement is the following:

**Theorem 6.1.** If Assumptions $[A] - [B]$ hold then a.s. as $(n, N) \to \infty$,

$$
d(\mu_S, \mu_c) \to 0.
$$

With stronger assumptions there is also convergence of the extremal eigenvalues of $S$.

**Theorem 6.2.** If Assumptions $[A] - [C]$ hold then for any $\varepsilon > 0$, a.s. for all $n \gg 1$,

$$
|\lambda_n - ((1 - \sqrt{c})^2| \leq \varepsilon \text{ and } |\lambda_1 - (1 + \sqrt{c})^2| \leq \varepsilon.
$$

In the sequel, we will use the notation $b_n = O_z(a_n)$ if there exist integers $k, \ell \geq 1$ such that $b_n = O((|z| + 1)^\ell (1 \wedge \Im(z))^{-k a_n})$. To prove Theorem 6.2, we will use a variant of the method developed in the previous chapter which avoids the Gaussian integration by parts.

### 6.2 Convergence of traces of resolvent matrices

#### 6.2.1 Resolvent identities

Let $A \in \mathcal{H}_n(\mathbb{R})$. For all $z \in \mathbb{C}_+ = \{z \in \mathbb{C} : \Im(z) > 0\}$, we define

$$
G_A(z) = (A - z)^{-1}.
$$

Since $A$ is symmetric, $G_A$ is also symmetric. We recall the resolvend identity

$$
G_B = G_A - G_A(B - A)G_B. \quad (6.2.1)
$$

Moreover for any $u, v \in \mathbb{R}^n$, the Sherman-Morrison formula implies that

$$
G_{A + uv^t} = G_A - \frac{G_Auv^tG_A}{1 + v^tG_Au}. \quad (6.2.2)
$$

The following lemma is an easy consequence Sherman-Morrison formula.

**Lemma 6.3.** We have

$$
G_{A + XX^t}XX^t = \frac{G_AXX^t}{1 + X^tG_AX}
$$

and

$$
\text{Tr}G^2_{A + XX^t}XX^t = \frac{X^tG_A^2X}{(1 + X^tG_AX)^2}.
$$

**Proof.** The first statement follows directly from (6.2.2):

$$
G_{A + XX^t}XX^t = G_AXX^t - \frac{G_AXX^tG_AXX^t}{1 + X^tG_AX} = G_AXX^t - \frac{(X^tG_AX)G_AXX^t}{1 + X^tG_AX}.
$$

For the second, we write

$$
G^2_{A + XX^t}XX^t = \left( G_A - \frac{G_AXX^tG_A}{1 + X^tG_AX} \right) \left( G_A - \frac{G_AXX^tG_A}{1 + X^tG_AX} \right) XX^t
$$

$$
= G_AXX^t - \frac{G^2_AXX^tG_AXX^t}{1 + X^tG_AX} - \frac{G_AXX^tG^2_AXX^t}{1 + X^tG_AX} + \frac{G_AXX^tG^2_AXX^t}{(1 + X^tG_AX)^2}.
$$

It remains to group terms as above and take the trace. \qed
6.2. CONVERGENCE OF TRACES OF RESOLVENT MATRICES

6.2.2 Concentration inequalities

We now set $G = G_S$ where $S$ is defined by (6.1.1) and

$$D(z) = \text{diag}(G_{11}(z), \ldots, G_{nn}(z)).$$

For integer $k \geq 1$, we now consider an arbitrary product of $G$ or $D$:

$$P(z) = \prod_{\ell=1}^{k} A_\ell(z)$$

where $A_\ell \in \{G, G^*, D, D^*\}$.

**Lemma 6.4.** Let $P$ be as above. For any $t > 0$, we have

$$P \left( \left| \frac{1}{n} \text{Tr} P(z) - \mathbb{E} \frac{1}{n} \text{Tr} P(z) \right| \geq t \right) \leq 2 \exp \left( -\frac{n^2 t^2 \text{Im}(z)^{2k}}{8k^2 N} \right).$$

**Proof.** For any $X \in \mathcal{H}_n(\mathbb{C})$, we denote by

$$P(X) = \prod_{\ell=1}^{k} A_\ell(X),$$

where $A_\ell(X) \in \{G_X, G_X^*, D_X, D_X^*\}$ (the dependency in $z$ is implicit). We write the telescopic sum

$$P(X) - P(Y) = \sum_{i=1}^{k} \left( \prod_{\ell=1}^{i-1} A_\ell(Y) \right) (A_i(X) - A_i(Y)) \left( \prod_{\ell=i+1}^{k} A_\ell(X) \right)$$

$$= \sum_{i=1}^{k} P_i(A_i(X) - A_i(Y))Q_i.$$

We find

$$|\text{Tr} P(X) - \text{Tr} P(Y)| \leq \sum_{i=1}^{k} |\text{Tr}(Q_i P_i(A_i(X) - A_i(Y)))|$$

We claim that for any $1 \leq i \leq k$,

$$|\text{Tr}(Q_i P_i(A_i(X) - A_i(Y)))| \leq 2 \text{Im}(z)^{-k} \text{rank}(X - Y). \quad (6.2.3)$$

Consider first the case where $A_i(X) \in \{G, G^*\}$. The resolvent identity (6.2.1) implies that

$$\text{rank}(P_i(A_i(X) - A_i(Y))Q_i) \leq \text{rank}(A_i(X) - A_i(Y)) \leq \text{rank}(X - Y).$$

Moreover, $\|Q_i P_i(A_i(X) - A_i(Y))\| \leq 2 \text{Im}(z)^{-k}$ and (6.2.3) follows.

We now deal with the case $A_i(X) \in \{D_X, D_X^*\}$. We set $B = G_X - G_Y$ if $A_i(X) = D_X$ and $B = G_X^* - G_Y^*$ if $A_i(X) = D_X^*$. Then, from (6.2.1) the matrix $B$ has rank $r \leq \text{rank}(X - Y)$ and norm bounded by $2 \text{Im}(z)^{-1}$. We perform the singular value decomposition of $B$:

$$B = \sum_{\alpha=1}^{r} s_\alpha u_\alpha v_\alpha^*,$$

with $\|u_\alpha\|_2 = \|v_\alpha\|_2 = 1$ and $0 \leq s_\alpha \leq 2 \text{Im}(z)^{-1}$. Also, since $\|Q_i P_i\| \leq \text{Im}(z)^{-k+1}$, we find that for any $1 \leq j \leq n_i$,

$$|(Q_i P_i)_{jj}| \leq 2 \text{Im}(z)^{-k} \sum_{\alpha=1}^{r} |e_j^\alpha u_\alpha| |e_j^\alpha v_\alpha|.$$
So finally, using Cauchy Schwartz inequality,

$$|\text{Tr}(Q_iP_i(A_i(X) - A_i(Y)))| \leq 2\text{Im}(z)^{-k} \sum_{j=1}^{r} \sum_{\alpha=1}^{n} |e_{j}^{t}u_{\alpha}| |e_{j}^{t}v_{\alpha}|$$

$$\leq 2\text{Im}(z)^{-k} \sum_{\alpha=1}^{n} \left(\sum_{j=1}^{n} |e_{j}^{t}u_{\alpha}|^2\right) \left(\sum_{j=1}^{n} |e_{j}^{t}v_{\alpha}|^2\right)$$

$$\leq 2\text{Im}(z)^{-k}\text{rank}(X - Y).$$

It proves (6.2.3). We get

$$|\text{Tr}P(X) - \text{Tr}P(Y)| \leq 2k\text{Im}(z)^{-k}\text{rank}(X - Y).$$

The conclusion follows by applying Azuma-Hoeffding’s inequality as in the proof of Theorem 3.17.

### 6.2.3 Variance bounds

The next result gives a better variance estimates in $n$ of $\frac{1}{n}\text{Tr}G(z)$ than Lemma 6.4 when Assumption (B) holds.

**Proposition 6.5.** Assume that Assumption (A) and (B) holds. Then,

$$\text{Var}\left(\frac{1}{n}\text{Tr}G(z)\right) = O_{z}\left(\varepsilon_{n}^{2}\right).$$

**Lemma 6.6.** For any real $(x_{j})_{1 \leq j \leq n}$, $(y_{j})_{1 \leq j \leq n}$, $y_j \geq 0$, $v > 0$,

$$|1 + \sum_{j=1}^{n} \frac{y_{j}}{x_{j} - iv}| \geq v.$$

**Proof.** Write $y_{j} = y p_{j}$ with $\sum_{j} p_{j} = 1$. The image of $x \mapsto \frac{1}{x - iv}$, $x \in \mathbb{R}$, is a circle $C$ of center $2i/v$ and radius $2/v$. It follows that the set of $z \in \mathbb{C}$ such that $z = \sum_{j=1}^{n} \frac{y_{j}}{x_{j} - iv}$ for some $(p_{j})$, $(x_{j})$ as above is a disc $D$ of center $2i/v$ and radius $2/v$. In particular $|1 + \sum_{j=1}^{n} \frac{y_{j}}{x_{j} - iv}|$ is lower bounded by $y$ times the distance of $-1/y$ to $D$. This distance is minimized for some $z \in C$ and we thus need to lower bound, for all real $x, y$

$$|1 + \frac{y}{x - iv}|^2 = \left(1 + \frac{y x}{x^2 + v^2}\right)^2 + \frac{y^2 v^2}{(x^2 + v^2)^2}.$$

In $y$, the infimum is reached for $y = -x$. It follows that

$$|1 + \frac{y}{x - iv}|^2 \geq \left(1 - \frac{x^2}{x^2 + v^2}\right)^2 + \frac{x^2 v^2}{(x^2 + v^2)^2} = v^2.$$

The conclusion follows.

**Lemma 6.7.** Assume that Assumption (A) and (B) holds. Let $\ell \in \mathbb{N}$, $A \in \mathcal{H}_n(\mathbb{C})$, $B \in \mathcal{M}_n(\mathbb{C})$ and $R = (A - z)^{-1}$, $z \in \mathbb{C}_+$. We have

$$E \left|\frac{1}{(1 + cY^{t}RY)^{\ell}} - \frac{1}{(1 + \varepsilon_{n} R)^{\ell}}\right|^2 = O_{z, \ell}(\varepsilon_{n}^2),$$

and

$$E \left|\frac{Y^{t}BY}{(1 + cY^{t}RY)^{\ell}} - \frac{\varepsilon_{n} R}{(1 + \varepsilon_{n} R)^{\ell}}\right|^2 \leq O_{z, \ell}(\|B\|^2 \varepsilon_{n}^{2}).$$
6.3. CONVERGENCE OF TRACES OF RESOLVENT MATRICES

Proof. We set \( f = \frac{1}{n} \text{Tr} R \) and \( g = \frac{1}{n} \text{Tr} B \). We have

\[
\left| \frac{1}{x^t} - \frac{1}{y^t} \right| \leq \frac{\ell|x - y|}{(|x| \wedge |y|)^{\ell - 1}(|x| \vee |y|)}.
\] (6.2.4)

By Lemma 6.6 we deduce that

\[
E \left[ \frac{1}{(1 + cY^t R)^t} - \frac{1}{(1 + cf)^t} \right]^2 \leq O_{\varepsilon, t}(|Y^t R - f|^2).
\]

The first statement follows. Similarly, we write

\[
\frac{Y^t B Y}{(1 + cY^t R)^t} - \frac{g}{(1 + cf)^t} = \frac{Y^t B Y - g}{(1 + cY^t R)^t} + g \left( \frac{1}{(1 + cY^t R)^t} - \frac{1}{(1 + cf)^t} \right).
\]

We deduce that

\[
\left| \frac{Y^t B Y}{(1 + cY^t R)^t} - \frac{1}{n} \text{Tr} B \right| \leq 2 \left( \frac{|Y^t B Y - g|^2}{(1 + cY^t R)^{2t}} + 2 |g|^2 \right)^{1/2} \left( \frac{1}{(1 + cY^t R)^t} - \frac{1}{(1 + cf)^t} \right)^{1/2}.
\]

Using Lemma 6.6 and the previous bound, it yields to the claimed expression. \( \square \)

Proof of Proposition 6.7. We have \( F(X_1, \ldots, X_N) = \frac{1}{n} \text{Tr} G \) for some function \( F : \mathbb{R}^{nN} \to \mathbb{R} \).

From Efron-Stein inequality,

\[
\text{Var} \left( \frac{1}{n} \text{Tr} R \right) \leq \frac{1}{2} \sum_{i=1}^{N} E |F(X) - F(X_{-i})|^2 = \frac{N}{2} E |F(X) - F(X_{-1})|^2
\]

where \( X = (X_1, \ldots, X_N) \) and \( X_{-i} = (X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_N) \) and \( X'_i \) is an independent copy of \( X_i \).

We set

\[ S_j = c \sum_{i \neq j} X_i X'_i \quad \text{and} \quad G_j = (S_j - zI)^{-1}. \]

Observe that \( G_j \) and \( X_j \) are independent. We denote by \( E_j \) the expectation with respect to \( X_j \).

By (6.2.2), we get

\[
\text{Var} \left( \frac{1}{n} \text{Tr} R \right) \leq \frac{N}{2} E \left| \frac{1}{n} \text{Tr} \left( \frac{cG_1 X_1 X'_1 G_1}{1 + cX'_1 G_1 X_1} \right) - \frac{1}{n} \text{Tr} \left( \frac{cG_1 X'_1 X'_1 G_1}{1 + cX'_1 G_1 X_1} \right) \right|^2
\]

\[
= \frac{c^2 N}{2n^2} E \left| X'_1 G_1^2 X_1 - X'_1 G_1^2 X_1 \right|^2
\]

\[
= \frac{c^2 N}{n^2} \text{Var}_1 \left( \frac{X'_1 G_1^2 X_1}{1 + cX'_1 G_1 X_1} \right),
\]

where \( \text{Var}_1 \) denotes the variance under \( E_1 \). It remains to apply Lemma 6.7 and recall that \( \|G_1\| \leq \text{Im}(z)^{-1} \). \( \square \)

6.3 Convergence of traces of resolvent matrices

We introduce the analytic functions

\[ f(z) = \frac{1}{n} \text{Tr} G(z) = \frac{1}{n} \text{Tr} D(z), \]
CHAPTER 6. COVARIANCE MATRICES

\[ g(z) = \frac{1}{n} \text{Tr}G(z)^2 = \frac{1}{n} \text{Tr}G(z)G(z)^t \]

and

\[ h(z) = \frac{1}{n} \text{Tr}D(z)^2. \]

In this section, we compute an asymptotic equivalent for \( \mathbb{E}f(z), \mathbb{E}g(z) \) and \( \mathbb{E}h(z) \).

**Proposition 6.8.** Under assumption (A)-(B), we have

\[ \mathbb{E}f(z) = f_c(z) + O_z(\varepsilon_n + n^{-1}). \]

where \( f_c(z) = \int \frac{1}{x-z} d\mu_c(x) \).

**Lemma 6.9.** The function \( f_c(z) = \int \frac{1}{x-z} d\mu_c(x) \) is analytic on \( \mathbb{C} \setminus [(1 - \sqrt{c})^2, (1 + \sqrt{c})^2] \) and satisfies

\[ czf_c(z) = 1 - c - \frac{1}{1 + cf_c(z)}. \]

**Proof.** \( \cdots \)

We then establish an approximated fixed point equation for \( \mathbb{E}f(z) \).

**Lemma 6.10.** For any \( z \in \mathbb{C}_+ \),

\[ cz\mathbb{E}f(z) = 1 - c - \frac{1}{1 + c\mathbb{E}f(z)} + O_z(\varepsilon_n + n^{-1}). \]

**Proof.** We write

\[ zG = -I + GS = -I + c \sum_{i=1}^N GX_iX_i^t. \]

We set

\[ S_j = c \sum_{i \neq j} X_iX_i^t \quad \text{and} \quad G_j = (S_j - zI)^{-1}. \]

We use Lemma 6.3:

\[ zG = -I + c \sum_{i=1}^N \frac{G_iX_iX_i^t}{1 + cX_i^tG_iX_i}. \]

Taking trace and expectation,

\[ z\mathbb{E}f(z) = -1 + \mathbb{E} \frac{\text{Tr}G_1X_1^tX_1^t}{1 + cX_1^tG_1X_1} = -1 + \mathbb{E} \frac{X_1^tG_1X_1}{1 + cX_1^tG_1X_1}. \]

So finally

\[ cz\mathbb{E}f(z) = 1 - c - \mathbb{E}(1 + cX_1^tG_1X_1)^{-1}. \]  \hspace{1cm} (6.3.1)

Lemma 6.7 and the independence of \( X_1 \) and \( G_1 \) yield to

\[ \mathbb{E} \frac{1}{1 + cX_1^tG_1X_1} = \frac{1}{1 + \frac{\varepsilon}{n} \text{Tr}G_1} + O_z(\varepsilon_n). \]

Using (6.2.4) and Lemma 6.6, we also get

\[ \left| \mathbb{E}(1 + \frac{c}{n} \text{Tr}G_1)^{-1} - (1 + c\mathbb{E}f(z))^{-1} \right| = O_z \left( \left| \mathbb{E} \left| \frac{1}{n} \text{Tr}G_1 - \frac{1}{n} \text{Tr}G \right| + \mathbb{E} \left| \frac{1}{n} \text{Tr}G - f(z) \right| \right) \].

By assumption (B) and Cauchy-Schwartz inequality, the first term is \( O(\Im(z)^{-2} \varepsilon_n) \). By Lemma 6.5, the second term is \( O(n^{-1} \Im(z)^{-2}) \). Finally, the last term is \( O_z(\varepsilon_n n^{-1/2}) \) by Proposition 6.5. It yields to the lemma. \( \Box \)
6.3. CONVERGENCE OF TRACES OF RESOLVENT MATRICES

The next lemma asserts that the quadratic equation satisfied by \( x = cf_\epsilon(z) \) is stable.

**Lemma 6.11.** Let \( x, y, z \in \mathbb{C}_+ \) such that \( zx^2 + x(z + c - 1) + c = 0 \) and \( zy^2 + y(z + c - 1) + c = \delta \) for some \( \delta \in \mathbb{C} \). Then

\[
|x - y| = O_\epsilon(|\delta|).
\]

**Proof.** We have \( y^2 + y(1 + (c - 1)z^{-1}) + cz^{-1} = \delta z^{-1} \) and

\[
\left( y - \frac{1}{2}(1 + (c - 1)z^{-1}) \right)^2 = \delta z^{-1} - cz^{-1} + \left( \frac{1}{2}(1 + (c - 1)z^{-1}) \right)^2.
\]

Subtracting the same expression with \( \delta = 0 \), we deduce that

\[
\delta z^{-1} = (y - x)(y + x + 1 + (c - 1)z^{-1}).
\]

Since, \( x + 1 + (c - 1)z^{-1} = -c/(zx) \),

\[
|x - y| = \frac{|\delta|}{|z||y - \frac{c}{zx}|}.
\]

It is sufficient to prove that for some \( \varepsilon > 0 \),

\[
\left| y - \frac{c}{zx} \right| \geq \varepsilon \frac{\Im(z)}{1 + |z|^2}.
\]

First, by assumption \( x = c \int_0^\infty \frac{1}{t-z} d\mu_c \). In particular, if \( z = u + iv \),

\[
 zx = c \int_0^\infty \frac{(u + iv)((t-u) + iv)}{(t-u)^2 + v^2} d\mu_c(t). \tag{6.3.2}
\]

Hence, \( zx \in \mathbb{C}_+ \),

\[
\Im(zx) = c \int_0^\infty \frac{vt}{(t-u)^2 + v^2} d\mu_c(t) \geq c \int_0^\infty \frac{(1 + \sqrt{\varepsilon})^2}{2u^2 + 2t^2 + v^2} d\mu_c(t) \geq \frac{cv}{2|z|^2 + 2(1 + \sqrt{\varepsilon})^4}.
\]

So finally, \(-1/(zx) \in \mathbb{C}_+ \) and for some \( \varepsilon > 0 \),

\[
\left| y - \frac{c}{zx} \right| \geq c \Im(-zx)^{-1} = \frac{\Im(zx)}{|zx|^2} \geq \varepsilon \Im(zx),
\]

where we have used that \( \Im(y) \geq 0 \) and \( |zx| = O(1) \) (which follows easily from (6.3.2)). \( \square \)

We can now complete the proof of Proposition 6.8.

**Proof of Proposition 6.8.** Let \( y = c\mathbb{E}f(z) \). By Lemma 6.10 we have \( zy^2 + y(z + c - 1) + c = O_\epsilon(n^{-\gamma/2}) \). It remains to use Lemma 6.11 and Lemma 6.9. \( \square \)

We now deal with \( \mathbb{E}g(z) \).

**Proposition 6.12.** Let

\[
g_c(z) = -\frac{f_c(z)}{cz + (zc f_c(z) + c - 1)^2} = \int \frac{1}{(x-z)^2} d\mu_c(x).
\]

Then \( g_c \) is analytic on \( \mathbb{C}\setminus[(1 - \sqrt{\varepsilon})^2, (1 + \sqrt{\varepsilon})^2] \). Under assumption [A] - [B] we have

\[
\mathbb{E}g(z) = g_c(z) + O_\epsilon(z_n + n^{-1/2}).
\]
Chapter 6. Covariance Matrices

Proof. We write

\[ G^2 = G(-z^{-1} I + cz^{-1} \sum_{i=1}^N G X_i X_i') = -z^{-1} G + cz^{-1} \sum_{i=1}^N G^2 X_i X_i'. \]

We take the trace and use Lemma 6.3, we find

\[ zg(z) = -f(z) + \frac{1}{N} \sum_{i=1}^N \frac{X_i' G^2 X_i}{(1 + c X_i' G_i X_i)^2}. \]

Taking expectation, we find

\[ z\mathbb{E}g(z) = -\mathbb{E}f(z) + \mathbb{E} \frac{X_1' G^2 X_1}{(1 + c X_1' G_1 X_1)^2}. \]

By Lemma 6.7, we find

\[ \mathbb{E} \frac{X_1' G^2 X_1}{(1 + c X_1' G_1 X_1)^2} = \frac{1}{n} \text{Tr} G^2_1 + O_z(\varepsilon_n). \]

By Lemma 6.4, \( \text{Var}(\frac{1}{n} \text{Tr} G^2) = O_z(\frac{1}{n}) \). Hence, arguing as in Lemma 6.10, we find

\[ cz\mathbb{E}g(z) = -f_c(z) - \frac{\mathbb{E}g(z)}{(1 + cf_c(z))^2} + O_z\left(\varepsilon_n + n^{-1/2}\right). \]

By Lemma 6.9 we may rewrite this expression as

\[ \mathbb{E}g(z)(cz + (czf_c(z) + c - 1)^2) = -f_c(z) + O_z\left(\varepsilon_n + n^{-1/2}\right). \]

We finally deal with \( \mathbb{E}h(z) \).

Proposition 6.13. Under assumption \( \text{(A)}-\text{(B)} \) we have

\[ \mathbb{E}h(z) = f_c^2(z) + O_z(\varepsilon_n + n^{-1}). \]

Proof. We have \( \mathbb{E}h(z) = \mathbb{E}G_{11}^2 \). By Proposition 6.8, \( \mathbb{E}G_{11} = f(z) = f_c(z) + O_z(\varepsilon_n) \), it is thus sufficient to check that

\[ \text{Var}G_{11} = O_z(\varepsilon_n^2). \]

We start with the identity

\[ zG_{11} = -1 + c \sum_{i=1}^N \frac{e_1' G_i X_i X_i' e_1}{1 + X_i' G_i X_i} \]

So that

\[ \text{Var}(G_{11}) = (c/z)^2 \text{Var} \left( \sum_{i=1}^N \frac{e_1' G_i X_i X_i' e_1}{1 + X_i' G_i X_i} \right). \]

\( \cdots \) FIXME : always the same computations...
6.4 Bound on matrix norm

Proposition 6.14. Under Assumptions (A)-(B) with \( \varepsilon_n = O(1/\sqrt{n}) \), we have
\[
\mathbb{E} \| S \| = O(\sqrt{\log n}).
\]

**Proof.** Since \( \mathbb{E} X_i X_i^* = I_n/n \), By symmetrization, we have
\[
\mathbb{E} \| S - I \| \leq 2 \mathbb{E} \left\| \sum_{i=1}^{N} \xi_n \frac{n}{N} X_i X_i^* \right\|,
\]
with \( \xi_n \) i.i.d. symmetric Bernoulli independent of \( X_1, \ldots, X_N \). From Rudelson’s inequality,
\[
\mathbb{E} \| S - I \| \leq C \sqrt{\log n} \sqrt{\mathbb{E} \left[ \sum_{i=1}^{N} \left( n/N \right)^2 \| X_i \|^2 \right]}.
\]

If \( \Lambda = \max_{1 \leq i \leq N} \| X_i \|^2 \), we find from Cauchy-Schwartz inequality and the triangle inequality,
\[
\mathbb{E} \| S - I \| \leq C' \sqrt{\log n} \sqrt{\mathbb{E} \| S - I \| + 1}.
\]

Assumption (B) with \( \varepsilon_n = O(1/\sqrt{n}) \) and \( A = I_n \) give that \( \mathbb{P}(\| X_i \|^2 \geq 1 + t) = O(1/(nt^2)) \). In particular, from the union bound,
\[
\mathbb{P}(\Lambda \geq 1 + t) = O(1/t^2).
\]
Consequently, \( \mathbb{E} \Lambda = O(1) \) and for some new constant \( C' \),
\[
\mathbb{E} \| S - I \| \leq C' \sqrt{\log n} \left( \sqrt{\mathbb{E} \| S - I \| + 1} \right).
\]

It is easy to check that if \( x^2 \leq a(x+1) \) then \( x \leq (\sqrt{2a}) \wedge (2a^2) \). We thus have proved that \( \mathbb{E} \| S \| \leq C'' \sqrt{\log n} \).

6.5 Convergence of edge eigenvalues

We introduce a new assumption on \( Y \),

(C”) \( Y \) has i.i.d. coordinates, \( \mathbb{E} |\sqrt{n} Y_i|^4 < C \) and for some \( q > 2 \), \( |\sqrt{n} Y_i| \leq n^{1/q} \) with probability 1.

We now establish the convergence in probability of extremal eigenvalues.

Proposition 6.15. Under Assumptions (A)-(C”), for any \( \varepsilon > 0 \), we have,
\[
\mathbb{P}(\| \lambda_n - ((1 - \sqrt{c})^2 + \varepsilon) \| \geq \varepsilon) = o(1) \quad \text{and} \quad \mathbb{P}(\| \lambda_1 - (1 + \sqrt{c})^2 \| \geq \varepsilon) = o(1).
\]

The passage from Proposition 6.15 to Theorem 6.2 is performed as in the previous chapter: we truncate and renormalize the entries of the vector \( Y \) which satisfies (C) (see Lemma 5.10). The almost sure convergence is obtained by an application of Talagrand’s concentration inequality.

Note that Proposition 6.8 and Lemma 6.4 imply that a.s.
\[
\limsup_{n \to \infty} \lambda_n \leq ((1 - \sqrt{c})^2) \quad \text{and} \quad \limsup_{n \to \infty} \lambda_1 \geq (1 + \sqrt{c})^2.
\]
We should thus proved the converse statement. We will follow the approach used in the previous chapter. We fix \( \kappa_n = \log n \). From Proposition 6.14

\[
\mathbb{P}(\|S\| > \kappa_n) = o(1).
\]  (6.5.1)

For any \( \varepsilon > 0 \), consider a smooth function \( \varphi : \mathbb{R}_+ \to [0, 1] \) with support

\[
I_\varepsilon = [0, (1 - \varepsilon)^2] \cup [(1 + \varepsilon)^2, \varepsilon, \kappa_n]
\]
such that \( \varphi(x) = 1 \) on \( I_{2\varepsilon} \).

Assume that we manage to prove that

\[
\mathbb{E} \left[ \frac{1}{n} \sum_{k=1}^n \varphi(\lambda_k) \right] = \mathbb{E} \int \varphi d\mu_M = o \left( \frac{1}{n} \right).
\]  (6.5.2)

Then, using \( 1(\lambda \in I_{2\varepsilon}) \leq \varphi(\lambda) \), we would deduce, that

\[
\mathbb{P}(\exists \lambda_k \in I_{2\varepsilon}) \leq \sum_{k=1}^n \mathbb{E} \varphi(\lambda_k) \leq n \mathbb{E} \int \varphi d\mu_M = o(1).
\]

Using (6.5.1), it would conclude the proof of Proposition 6.15.

The main technical lemma is the following.

**Lemma 6.16.** Under Assumptions (A)-(C), there exists \( \rho > 0 \) and an analytic function \( \tau \) on \( \mathbb{C} \setminus [(1 - \sqrt{\varepsilon})^2, (1 + \sqrt{\varepsilon})^2] \) such that for any \( z \in \mathbb{C}_+ \),

\[
\mathbb{E} f(z) = f_c(z) + \frac{\tau(z)}{n} + O_z \left( \frac{1}{n^{1+\rho}} \right).
\]

Before proving Lemma 6.16 let us give the proof of Proposition 6.15.

**Proof of Proposition 6.15.** We consider the above smooth function \( \varphi \) with support \( I_{\varepsilon} \). The term \( \delta(z)/n^{1+\gamma} = O_z(1/n^{1+\gamma}) \) in Lemma 6.16 is of the form \( \delta(z) = O((1 + |z|)^{\ell}(1 \land \text{Im}(z))^{-k}) \) for some integers \( k, \ell \geq 1 \).

Let \( K_\varepsilon = \{ z \in \mathbb{C}_+ : \Re(z) \in I_{\varepsilon} \} \). We may apply Lemma 5.5 with \( k \) as above. Since \( \partial \Phi(x + iy) \) has support in \( K_\varepsilon \), we get

\[
\mathbb{E} \int \varphi d\mu_M = \int_{K_\varepsilon} \partial \Phi(x + iy) \mathbb{E} f(x + iy) dxdy
\]

\[
= \int_{K_\varepsilon} \partial \Phi(x + iy) \left( f_c(x + iy) - \frac{\tau(x + iy)}{n} + \frac{\delta(x + iy)}{n^{1+\rho}} \right) dxdy
\]

\[
= \int_{K_\varepsilon} \partial \Phi(x + iy) \left( f_c(x + iy) - \frac{\tau(x + iy)}{n} \right) dxdy
\]

\[
+ \frac{1}{n} \Re \int_{K_\varepsilon} \partial \Phi(x + iy) \frac{\delta(x + iy)}{n^{1+\rho}} dxdy
\]

Now, notice that \( f_c(z) - \tau(z)/n \) is analytic on an open neighborhood of \( K_\varepsilon \). In particular, \( \partial(f_c - \tau/n) = 0 \) on this neighbourhood. Hence, by integration by part, the first integral of the above expression is 0. Also, from (5.2.10), for some \( C > 0 \),

\[
\partial \Phi(x + iy) \delta(x + iy) = O \left( 1(y \leq C)(1 + |x|^{\ell}) \right).
\]

We integrate over \( K_\varepsilon \) and recall that \( \kappa_n = n^{o(1)} \). We have proved that

\[
\mathbb{E} \int \varphi d\mu_M = o(n^{-1}).
\]

It concludes the proof of Proposition 6.15. \( \square \)
Lemma 6.17. Set \( f_1(z) = \frac{1}{n} \Tr G_1(z) \). Under Assumptions (A), (B),
\[
\mathbb{E} f(z) - \mathbb{E} f_1(z) = -\frac{1}{n} \frac{c g_c}{1 + c f_c} + O_z \left( \frac{\varepsilon_n}{n} \right) + O_z (1 / n^{3/2}) .
\]

Proof. By [6.2.2],
\[
G = G_1 - c \frac{G_1 X_1^t X_1 G_1}{1 + c X_1^t G_1 X_1} .
\]
Hence
\[
f(z) - f_1(z) = -\frac{c}{n} \Tr G_1 X_1^t G_1 = -\frac{c}{n} \frac{X_1^t G_1 X_1}{1 + c X_1^t G_1 X_1} .
\]
By Lemma 6.7, we find
\[
\mathbb{E} X_1^t G_1^2 X_1 = \frac{g_1}{1 + c f_1} + O_z (\varepsilon_n) ,
\]
where \( g_1 = \frac{1}{n} \Tr G_1^2 \). We may then write
\[
\frac{g_1}{1 + c f_1} = \mathbb{E} g_1 + \mathbb{E} \frac{g_1 - g_1}{1 + c f_1} - c \mathbb{E} g_1 \mathbb{E} \frac{f_1 - f_1}{(1 + c f_1) (1 + c f_1)} .
\]
It remains to apply Propositions 6.8-6.12 and Lemmas 6.5-6.6.

The next lemma establishes a convergence rate for random quadratic forms.

Lemma 6.18. Under Assumption (A) - (C'), that for any \( A \in M_n(\mathbb{C}) \),
\[
\mathbb{E} \left[ Y^t A Y - \frac{1}{n} \Tr A \right]^4 = O \left( \| A \|^4 n^{4/(q^A) - 3} \right)
\]

Proof. We can assume without loss of generality that \( A \in M_n(\mathbb{R}) \). We set \( Z_i = Y_i^2 - 1/n \) and write
\[
\mathbb{E} \left[ Y^t A Y - \frac{1}{n} \Tr A \right]^4 \leq 8 \mathbb{E} \left[ \sum_{i=1} A_{ii} Z_i \right]^4 + 8 \mathbb{E} \left[ \sum_{i \neq j} A_{ij} Y_i Y_j \right]^4 .
\]

We first treat the first term, we have
\[
\mathbb{E} \left[ \sum_{i=1} A_{ii} Z_i \right]^4 = \sum_{i} A_{ii}^4 \mathbb{E} (Z_i)^4 + \sum_{i \neq j} A_{ii}^2 A_{jj}^2 \mathbb{E} (Z_i)^2 \mathbb{E} (Z_j)^2 .
\]
Since \( |A_{ii}| \leq \| A \|, \| Z_i \| = O(n^{-2}) \) and \( |Z_i| = O(n^{2/q - 1}) \), we have \( \mathbb{E} (Z_i)^4 = O(n^{4/q - 4}) \) and
\[
\mathbb{E} \left\| \sum_{i=1} A_{ii} Z_i \right\|^4 = O(\| A \|^4 n^{4/q - 3}) .
\]
Similarly,
\[
\sum_{i \neq j} A_{ii}^2 A_{jj}^2 \mathbb{E} (Z_i)^2 \mathbb{E} (Z_j)^2 = O(\| A \|^4 n^{-2}) .
\]
Hence, the first term is \( [6.5.3] \) is bounded by \( O(\| A \|^4 n^{4/(q^A) - 3}) \).

For the second term of \( [6.5.3] \), up to replacing \( A \) by \( A = A - \text{diag}(A_{11}, \ldots, A_{nn}) \), we may simply assume that \( A_{ii} = 0 \) (indeed \( \| A' \| \leq \| A \| + \max_i \| A_{ii} \| \leq 2 \| A \| \)). We need expand
\[
\mathbb{E} \left( \sum_{i,j} A_{ij} Y_i Y_j \right)^4 = \sum_{\ell \leq \ell' \leq \ell''} \sum_{i,j} 4 A_{i\ell} A_{\ell'j} \mathbb{E} Y_{i \ell} Y_{\ell'j} \mathbb{E} Y_{\ell''j} Y_{j \ell} = \sum_{\ell \leq \ell' \leq \ell''} \prod_{a=1}^k A_{i\ell} A_{\ell'j} \prod_{a=1}^k \mathbb{E} Y_{s_a}^{m_a} ,
\]
where \((s_a), 1 \leq a \leq k\) is the set of distinct indices in \((i_t, j_t), 1 \leq t \leq 4\) and \(m_a\) is the number of occurrences of \(s_a\) in \((i_t, j_t), 1 \leq t \leq 4\). Since \(\mathbb{E} s_a = 0\), the above product is 0 unless \(m_a \geq 2\) for any \(1 \leq a \leq k\). It implies that \(1 \leq k \leq 4\) and that the ordered vector \((m_a)_{1 \leq a \leq k}\) is one of the following: \((2, 2, 2, 2), (3, 3, 2), (4, 2, 2), (4, 4), (5, 3), (6, 2)\) (the case \(k = 1\) and \(m_1 = 8\) does not contribute since \(A_{ii} = 0\)). By assumption, we have the bound

\[
\mathbb{E} Y_i^n = O\left(n^{-\frac{(p-4)\epsilon}{9} - \frac{2}{9}}\right). \tag{6.5.5}
\]

We deduce that

\[
\prod_{a=1}^{k} \mathbb{E} Y_{sa}^{m_a} = O(n^{-\frac{2}{3} + 1(k=2)}).
\]

We now deal with the control of \(\prod_{a=1}^{4} A_{ij}^{s_a}\). Let us start with the case \(k = 2\). We use twice the bound \(|A_{ij}| \leq \|A\|\). We get a contribution of order

\[
\left| \sum_{i,j} \prod_{a=1}^{4} A_{ij}^{s_a} \right| = O(\|A\|^{2n-4+2/q} \sum_{i,j} |A_{ij}|^2 + |A_{ij}A_{ji}|)
\]

From Cauchy-Schwartz inequality,

\[
\sum_{i,j} |A_{ij}A_{ji}| \leq \sum_{i,j} |A_{ij}|^2 = \text{Tr}(AA^*) \leq n\|A\|^2.
\]

So finally,

\[
\left| \sum_{i,j} \prod_{a=1}^{4} A_{ij}^{s_a} \right| = O(\|A\|^{4n-3+2/q}).
\]

For the terms such that \(k = 3\), we use once the bound \(|A_{ij}| \leq \|A\|\). It remains an expression of the form \(A_{ij}B_{jl}C_{kl}\) or \(A_{ij}B_{jk}C_{ki}\), where \(B\) and \(C\) are equal \(A\) or \(A^*\). In the first case, we use the bound

\[
\sum_{i,j,k} |A_{ij}B_{jl}C_{kl}| \leq \sum_{i,j} |A_{ij}| |B_{jl}| \sqrt{n} \left| \sum_{k} |C_{ik}|^2 \right| = n^{3/2} \|A\|^3.
\]

In the second case,

\[
\sum_{i,j,k} |A_{ij}B_{jk}C_{ki}| \leq \sum_{i,j} |A_{ij}| \left( \sum_{k} |B_{jk}|^2 \right)^{1/2} \left( \sum_{k} |C_{ik}|^2 \right)^{1/2} \leq \|A\|^2 \sum_{i,j} |A_{ij}| \leq n^{3/2} \|A\|^3.
\]

So finally,

\[
\left| \sum_{i,j,k} \prod_{a=1}^{4} A_{ij}^{s_a} \right| = O(\|A\|^{4n-4+3/2}).
\]

It remains to deal with \(k = 4\) in (6.5.4). By assumption \(\mathbb{E} Y_i = 1/n\). Each index appears exactly twice: there are expressions of the form \(A_{ij}B_{jl}C_{kl}D_{ik}\) and \(A_{ij}B_{jk}C_{kl}D_{ti}\) where \(B, C, D\) are equal \(A\) or \(A^*\). In the first case,

\[
\sum_{i,j,k,l} A_{ij}B_{jl}C_{kl}D_{ik} = \sum_{i,j} A_{ij}B_{jl} \sum_{k,l} C_{kl}D_{ik} = \text{Tr}(AB)\text{Tr}(CD),
\]

whose absolute value is bounded by \(n^2 \|A\|^2\). In the second case,

\[
\sum_{i,j,k,l} A_{ij}B_{jk}C_{kl}D_{ti} = \text{Tr}(ABCD),
\]

whose absolute value is bounded by \(n\|A\|^4\). We thus have checked that (6.5.4) is at most of order \(\|A\|^4 n^{-2}\).
6.5. CONVERGENCE OF EDGE EIGENVALUES

The next lemma computes the first moments of a random quadratic form.

**Lemma 6.19.** Under Assumptions (A)-(C'), we set \( \beta = n^2 E Y_1^4 = O(1) \). If \( A \in M_n(\mathbb{C}) \) and \( D = \text{diag}(A_{11}, \ldots, A_{nn}) \), we have

\[
E \left( Y^t A Y - \frac{1}{n} \text{Tr} A \right)^2 = \frac{1}{n^2} (\text{Tr} A^2 + \text{Tr} AA^t + (\beta - 2) \text{Tr} D^2),
\]

and

\[
E \left( Y^t A Y - \frac{1}{n} \text{Tr} A \right)^3 = O(\|A\|^3 n^{-2 + 2/q}).
\]

**Proof.** We expand brutally

\[
E \left( Y^t A Y - \frac{1}{n} \text{Tr} A \right)^2 = \sum_{i,j} \left( Y_i A_{ij} Y_j - E Y_i A_{ij} Y_j \right)^2
\]

where in the sum, the summand is zero unless one the two disjoint cases holds:

(i) \( i_1 = i_2 \) and \( j_1 = j_2 \)

(ii) \( i_1 = j_1, i_2 = j_2 \) and \( i_1 \neq j_1 \).

Let us start with case (i). We have

\[
(i) = \frac{1}{n^2} \sum_{i \neq j} A_{ij}^2 + \left( E Y_1^4 - \frac{1}{n^2} \right) \sum_i A_{ii}^2
\]

\[
= \frac{1}{n^2} \text{Tr} AA^t + \left( \frac{\beta}{n^2} - \frac{1}{n^2} \right) \text{Tr} D^2.
\]

Similarly,

\[
(ii) = \left( E Y_1^2 Y_2^2 \right) \sum_{i \neq j} A_{ij} A_{ji}
\]

\[
= \frac{1}{n^2} (\text{Tr} A^2 - \text{Tr} D^2),
\]

We now compute

\[
E \left( Y^t A Y - \frac{1}{n} \text{Tr} A \right)^3 = \sum_{i_1,j_1} \prod_{\ell=1}^3 A_{i_\ell,j_\ell} E \prod_{\ell=1}^3 \left( Y_{i_\ell} Y_{j_\ell} - E Y_{i_\ell} Y_{j_\ell} \right)
\]

\[
= \sum_{k=1}^3 S_k,
\]

where in \( S_k \), the sum is over \( k \) distinct values of \( i_{\ell}, j_{\ell} \) each appearing at least twice. If \( k = 1 \), we have from \((6.5.5)\)

\[
S_1 = \sum_i A_{ii}^3 E \left( Y_i^2 - \frac{1}{n} \right)^3 = O(\|A\|^3 n^{-2 + 2/q}).
\]
If \( k = 2 \), each distinct index appears 2, 3 or 4 times. From (6.5.5), the terms of the form \( A_{i\ell}A_{ij}A_{j'\ell} \) with \( i', j', i, j \) different from \( \ell \) are equal to 0 because \( \mathbb{E}(Y_\ell^2 - \mathbb{E}Y_\ell^2) = 0 \). Hence, only remains terms of the form \( A_{ij}B_{jk}C_{ki} \) where \( B, C \) are equal to \( A, A^t \) or \( D \). However,

\[
\sum_{i,j,k} A_{ij}B_{jk}C_{ki} = \text{Tr}(ABC) \leq n||A||^3.
\]

We deduce from (6.5.5) that

\[
S_2 = O(||A||^2n^{-2}).
\]

It remains to deal with \( S_3 \). Each distinct index appears exactly twice. Moreover we can restrict to sums such that \( i_\ell \neq j_\ell \) for any \( 1 \leq \ell \leq 3 \) (because \( \mathbb{E}(Y_\ell^2 - \mathbb{E}Y_\ell^2) = 0 \)). In particular, we find that \( \mathbb{E}Y_{i,j}_1 = 0 \) and \( \mathbb{E}Y_{i,j}_3 = 0 \) (since \( i_3 \neq j_3, i_3 \) is equal to exactly one of the index \( (i_1, i_2, j_1, j_2) \)). Expanding, we obtain

\[
S_3 = 0.
\]

\[\square\]

**Proof of Lemma 6.16** We set \( f_1 = \frac{1}{n} \text{Tr} G_1 \). By Lemma 6.19 Assumption (C’) implies that (B) holds with \( \varepsilon_n = O(n^{-1/2}) \). We start from (6.3.1)

\[
z c \mathbb{E} f(z) = 1 - c - \mathbb{E}(1 + cX_1^t G_1 X_1)^{-1}.
\]

We perform a Taylor expansion of \((1 + cX_1^t G_1 X_1)^{-1}\)

\[
\frac{1}{1 + cX_1^t G_1 X_1} = \sum_{\ell=0}^3 (-c)^\ell \frac{(X_1^t G_1 X_1 - f_1)^\ell}{(1 + cf_1)^{\ell+1}} + c^4 \frac{(X_1^t G_1 X_1 - f_1)^4}{(1 + cf_1)^4(1 + cX_1^t G_1 X_1)}.
\]

By Lemma 6.6 and Lemma 6.18 we have

\[
\mathbb{E}_1 \left| \frac{(X_1^t G_1 X_1 - f_1)^4}{(1 + cf_1)^3(1 + cX_1^t G_1 X_1)} \right| = O_z \left( \mathbb{E}_1 |X_1^t G_1 X_1 - f_1|^4 \right) = O_z \left( n^{4/(q \wedge 4) - 3} \right).
\]

We finally write,

\[
\mathbb{E}_1 \frac{1}{1 + cX_1^t G_1 X_1} = \frac{1}{1 + cf_1} + c^2 \mathbb{E}_1 (X_1^t G_1 X_1 - f_1)^2 \left( \frac{1 + cf_1}{1 + cf_1} \right)^3
\]

\[
- c^3 \mathbb{E}_1 (X_1^t G_1 X_1 - f_1)^3 \left( \frac{1 + cf_1}{1 + cf_1} \right)^4 + c^4 \mathbb{E}_1 \frac{(X_1^t G_1 X_1 - f_1)^4}{(1 + cf_1)^4(1 + cX_1^t G_1 X_1)}
\]

\[
= \frac{1}{1 + cf_1} + c^2 \mathbb{E}_1 \left( X_1^t G_1 X_1 - f_1 \right)^2 \left( \frac{1 + cf_1}{1 + cf_1} \right)^3 + O_z \left( n^{4/(q \wedge 4) - 3} \right),
\]

where we have applied Lemma 6.19 for the third term. It remains to apply Lemma 6.19 for the second term

\[
\mathbb{E}_1 \frac{1}{1 + cX_1^t G_1 X_1} = \frac{1}{1 + cf_1} + \frac{c^2 2g_1 + (\beta - 2)h_1(z)}{n} + O_z \left( n^{4/(q \wedge 4) - 3} \right),
\]

where \( h_1 = \frac{1}{n} \text{Tr}(D_1^2) \) and \( g_1 = \frac{1}{n} \text{Tr}(G_1^2) = \frac{1}{n} \text{Tr}(G_1 D_1) \). From (6.2.3), we find that

\[
|h_1(z) - h(z)| = O_z(1/n) \quad \text{and} \quad |g_1(z) - g(z)| = O_z(1/n).
\]

So finally, we may apply Propositions 6.12 and 6.13 and use Lemma 6.17 in (6.3.1) it leads to

\[
z c \mathbb{E} f(z) = 1 - c - \frac{1}{1 + c \mathbb{E} f(z)}
\]

\[
+ \frac{1}{n} \left( - \frac{c g_c}{(1 + cf_c)^3} - \frac{c^2 2g_c + (\beta - 2)h_c(z)}{(1 + cf_c)^3} \right) + O_z \left( n^{-3/2} + n^{4/(q \wedge 4) - 3} \right)
\]

\[
= 1 - c - \frac{1}{1 + c \mathbb{E} f(z)} + \frac{\varphi(z)}{n} + O_z \left( n^{1 - \rho} \right),
\]

where \( \varphi(z) \) is a polynomial of degree \( \leq 3 \) in \( z \).
where $\varphi(z)$ is analytic outside $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$ and where $\rho = 2 - 4/(q \wedge (8/3)) > 0$. We may rewrite this last equality as

$$z c^2 (\mathbb{E}f(z))^2 + c \mathbb{E}f(z)(z + c - 1 - \frac{\varphi(z)}{n}) + c - \frac{\varphi(z)}{n} = O_z(n^{-1-\rho}). \tag{6.5.7}$$

We may solve this quadratic equation. By Lemma 6.11 we have $\mathbb{E}f(z) = f_c(z) + \delta(z)/n$ with $\delta(z) = O_z(1)$. Putting this into (6.5.7) and keeping terms of order $O_z(n^{-1-\rho})$ give a linear equation satisfied by $\delta(z)$ in terms of $f_c$, $\varphi$ and $O_z(n^{-1-\rho})$. This concludes the proof of the lemma.
Bibliography


